3

LETI's SCIENTIFIC SERVICES

COLECCIÓN
LETITRAIN

Lectura crítica de un artículo científico

leti laboratorios

www.leti.com/zonaprofesional

COLECCIÓN
LETITRAIN

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado?

(1ª parte)

José Luis Mateos, PhD

"... El curso tradicional de lectura de la literatura médica consiste en: '¡Aquí tiene The New England Journal of Medicine'; *léala!*

Este método es análogo al de aprender a nadar por el método de la inmersión total.

Desde luego, ciertas personas pueden aprender a nadar de esa forma, pero algunas se ahogan y muchas le toman miedo al agua. Leer solo los resúmenes de los artículos médicos es como tener miedo al agua..."

(Riegelman RK, Hirsch. En: *Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica.* Biblioteca de la Organización Panamericana de la Salud. Little, Brown & Co. [Eds.] 1992. Boston, MA, USA)

© 2015, Laboratorios Leti
Realización editorial, Lexic S.L.
Passeig del Ferrocarril, 339, 2º 4ª - 08860 Castelldefels (Barcelona), España.

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado?

(1ª parte)

ÍNDICE

Índice

Introducción 1

1. La importancia de hacer una lectura crítica 5

1.1. ¿Por qué y cómo comenzar a leer un artículo científico? 11

1.1.1. ¿Por qué y cómo comenzar a leer un artículo científico? 13

1.1.2. El proceso de la medicina basada en la evidencia 18

1.2. Estructura básica de un artículo científico 26

1.2.1. Parte conceptual: introducción y objetivos .. 32

1.2.2. Parte de diseño: material y métodos 34

1.2.3. Parte empírica: resultados 38

1.2.4. Parte interpretativa: discusión y conclusiones 40

1.3. Los falsos mitos en la publicación de artículos 47

1.3.1. Mito 1. Si un artículo ha llegado a ser publicado es porque es válido 48

1.3.2. Mito 2. El prestigio profesional o académico de un autor es garantía de la calidad de un artículo 51
1.3.3. Mito 3. El antecedente de la publicación de un excelente trabajo en el pasado respalda la calidad de la producción actual 52
1.3.4. Mito 4. El investigador por la naturaleza objetiva de su trabajo está libre de prejuicios y de sesgos 56
1.4. ¿Qué es la lectura crítica de un artículo y qué se persigue? 60
1.5. Fases de la lectura crítica 71
1.6. Niveles de evidencia 78
1.7. Instrumentos para la lectura crítica 87

2. Criterios generales para la lectura crítica de artículos 97
2.1. Evaluación preliminar 99
2.2. Las preguntas más frecuentes 107
2.2.1. ¿Cuál es la pregunta de investigación? 109
2.2.2. ¿Se trata de un estudio original? 110
2.2.3. ¿Por qué se hizo el estudio? 112
2.2.4. ¿Qué tipo de estudio llevaron a cabo los investigadores? 115

2.2.5. ¿Tiene el estudio un diseño adecuado a su objetivo? 140
2.2.6. ¿Se han evitado o minimizado los sesgos? 155
2.2.7. ¿Es adecuada la población estudiada? 175
2.2.8. ¿Son adecuadas las características metodológicas? 187
2.3. El cumplimiento de la cuestión estadística 197
2.3.1. Objetivos de la cuestión estadística 200
2.3.2. Pruebas de significación estadística 211
2.3.3. Selección de las pruebas estadísticas adecuadas 219
2.3.4. Significación estadística e importancia clínica: valoración de la causalidad 229

3. Conclusiones 241

4. Bibliografía 249

5. Notas 261

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado?

(1ª parte)

INTRODUCCIÓN

La investigación clínica ha de entenderse en un contexto social y global en el que el clínico desempeña un papel principal, tratando de buscar respuesta a todas las cuestiones que se le presentan durante el ejercicio de su actividad. No hay duda alguna de que el desarrollo que ha experimentado la medicina en las últimas décadas se debe, sobre todo, al empuje que ha tenido la investigación clínica y a la correcta difusión del conocimiento de los resultados propios de estos avances médicos.

Los conocimientos sobre un tema o un campo de investigación clínica se adquieren básicamente a través de la literatura que se genera a partir de dicha investigación. Sin embargo, la búsqueda de bibliografía, la identificación de documentos o trabajos relevantes y la lectura prolongada y ordenada de revistas de nuestra especialidad no conducen a ninguna parte si no hay una lectura crítica de los artículos científicos.

Dirigido fundamentalmente a los profesionales sanitarios, en este libro se definen los criterios de validez, importancia y aplicabilidad, que son los pilares básicos para la lectura crítica de los trabajos científicos que la literatura médica nos ofrece actualmente. Así, en una primera parte, se revelará la importancia de leer críticamente un artículo, para pasar después a estudiar las normas o criterios generales para evaluar críticamente la literatura, tratándose con un especial cuidado las preguntas más frecuentes que ha de satisfacer la publicación de cualquier estudio, así como a la cuestión estadística, que

aún hoy resulta un tema ciertamente farragoso para el no iniciado; y luego, de forma individualizada, se presentarán los criterios para evaluar los principales tipos de diseño y de artículos de investigación clínica. La lectura crítica y sistemática de la literatura biomédica nos va a ayudar a conocer bien la literatura, lo que coloquialmente llamamos *separar el grano de la paja*, tanto en la preparación de nuestros trabajos como en la revisión o evaluación que podamos hacer de otros trabajos para alguna revista.

Lectura crítica de un artículo científico
¿Cómo analizar un artículo publicado?
(1ª parte)

1

LA IMPORTANCIA DE HACER UNA LECTURA CRÍTICA

1. LA IMPORTANCIA DE HACER UNA LECTURA CRÍTICA

La medicina, cuyo objetivo inmediato es el hombre en la enfermedad, es un arte que se sitúa en la relación directa entre médico y paciente. Goza de la profundidad del conocimiento científico que le otorgan precisión y certeza a la hora de tomar una decisión clínica. El proceso de la decisión clínica va a llevar al médico a tomar una decisión terapéutica y a la observación final del resultado del tratamiento. Dicho así, puede parecernos algo sencillo, aunque nada más lejos de la realidad. En el razonamiento clínico, interviene un proceso complejo en el que los datos recogidos de cada paciente, su manejo particular y su posterior integración en un nivel superior –*cuerpo de conocimiento*– de la enfermedad, son los pilares elementales para el buen hacer en la praxis médica.

La toma de decisiones para el médico ha de empezar por la recogida de datos del enfermo a partir del contacto directo entre médico y paciente. En este contexto, la mayoría de las veces que el médico ve a un paciente va a necesitar infor-

mación sobre algún elemento de su diagnóstico, pronóstico o tratamiento. Esta información que, en muchas ocasiones, es clara y está al alcance de la mano, en otras muchas, no será tan obvia y habremos de localizarla a partir de la evidencia externa. Para muchos profesionales de la salud, el esfuerzo requerido para responder a estas cuestiones es tan importante, pero su tiempo tan limitado, que la necesidad de información nunca se ve completada, por lo que incluso pueden quedar bloqueados en este proceso[1].

Frecuentemente, se tiende a sacralizar la *letra impresa*, es decir, se da por válida y relevante cualquier información que venga avalada por los revisores –*peer review*– de una revista prestigiosa, o no tan prestigiosa[2]. Ya sea de investigación básica, clínica o epidemiológica, dar por válido todo lo que caiga en nuestras manos puede suponer cambios en la toma de decisiones clínicas, cualquiera que sea su especialidad. Es obvio que no es una actitud muy inteligente dudar de todo lo que se publica en medicina, aunque es todavía peor aceptarlo todo sin evaluación crítica alguna.

Cuando nos dispongamos a leer un artículo científico –artículo original, revisión, metaanálisis– con el objetivo de incorporar su contenido a nuestro cuerpo de conocimiento, es necesario plantearse una duda razonable y mantener el control sistemático sobre nuestra lectura a través de una serie de pasos.

El manejo eficiente de la lectura de un artículo científico requiere de una sistemática concreta:

- Aclarar las razones para la lectura de un asunto.
- Especificar la información necesaria.
- Identificar los artículos relevantes del asunto que nos ocupa.
- Lectura crítica de los artículos.

La lectura crítica es una técnica que ofrece la oportunidad de aumentar la efectividad de nuestra lectura, adquiriendo con prontitud las habilidades necesarias para excluir aquellos artículos que se juzguen de mala calidad y aceptar solo aquellos con la suficiente calidad metodológica. Debemos utilizar, en la medida de lo posible, una serie de criterios que nos permitan hacer una evaluación crítica de lo que leemos en cada momento. Es decir, hemos de ser capaces de **verificar la validez** –aproximación a la verdad– **de los resultados o las evidencias presentadas**, de decidir si los **resultados son importantes y de utilidad clínica** –de aplicabilidad clínica en el modo de actuar con los pacientes–. Esto va a permitir engrosar el *cuerpo del conocimiento* del lector, de forma que los conocimientos evaluados tras la lectura puedan eventualmente incorporarse al cuidado de los pacientes y a la mejor calidad de la asistencia.

No en vano, un artículo científico se publica para obtener información utilizable, aplicable y basada en la evidencia que permita mejorar la práctica clínica con los pacientes.

En cuanto a los resultados, **los artículos científicos publicados deben ser evaluados en tres aspectos fundamentales**[3].

- **¿Podemos confiar en los resultados?** Dicho de otro modo, ¿son válidos estos resultados? Es decir, aprenderemos a juzgar la validez metodológica del artículo. Se trata de ver la calidad y el grado de adecuación del diseño del que se pretenden extraer las conclusiones. Como veremos a lo largo de los contenidos de este libro, los criterios de validez son diferentes según la pregunta de investigación (tratamiento, diagnóstico, pronóstico, evaluación económica).
- **¿Cuáles son los resultados?** Aquí aprenderemos a analizar la relevancia del resultado obtenido, respondiendo a preguntas como ¿la intervención experimental frente al control muestra algún efecto?, ¿cómo se mide el efecto?, ¿son precisos los resultados?
- **¿Puedo extrapolar los resultados obtenidos a mi práctica habitual?** Este es el objetivo final de la lectura crítica de un artículo, y quizá el más importante, por tratarse del de mayor pragmatismo. ¿Podré aplicar lo extraído de esta investigación en la atención médica de mis pacientes?

Y tras esta breve introducción, entramos ya en materia con la intención de contagiarle por su lectura con el mismo entusiasmo que tiene el que se dispone a escribirlo. ¡Ánimo, y a por ellos, que son pocos y cobardes...!

1.1. ¿Por qué y cómo comenzar a leer un artículo científico?

En la actualidad existen cerca de 25.000 revistas médicas y se estima que anualmente se publican más de 6.000 artículos con *abstract* (resumen) en las revistas de biomedicina de mayor índice de impacto. Ante estos datos, un médico generalista necesitaría leer un promedio de unos 17 artículos diarios para mantenerse a la vanguardia de su área de influencia profesional[4,5]. Por mencionar un ejemplo, solo para actualizarse en el área de las alergias, si se anima a buscar en PUBMED todo lo publicado en 2013 en inglés tras introducir el término "allergy", observará que aparecen un total de 13.458 artículos publicados. Es decir, cualquier alergólogo del mundo habría necesitado leer un promedio de algo más de 36 publicaciones diarias para estar al corriente de lo más importante que se publicó sobre alergias en general en 2013.

Así es que para que un alergólogo pudiera leer todo lo que se publica sobre "alergia" en un año necesitaría unas 18 horas diarias durante los 365 días del año. ¿Alguien se atreve?

Nada más lejos de la realidad, si tenemos en cuenta que el tiempo disponible para que el profesional pueda acceder a la literatura científica es cada vez menor, por tener que dedicarse a atender otras tareas ya sea de carácter institucional o privadas.

La literatura biomédica se expande a un ritmo del 6-7% anual. Por tanto, su número puede duplicarse cada 10-15 años, y se incrementa 10 veces cada 35-50 años6. Pero sigamos con nuestro ejemplo. La tendencia que ha experimentado la aparición de artículos científicos en inglés en alergología es significativamente exponencial. Tras elegir al azar el número de artículos publicados entre los años 2000 y 2012, se observa un incremento del 42%. Ante un volumen de publicaciones de tal calibre, resulta imposible que un alergólogo, o cualquier otro profesional de la salud pueda estar al día en todos los temas relacionados con la práctica clínica diaria. Esto puede llegar a generar un caos que complica la tarea del médico y lo aleja de la posibilidad de actualizar su cuerpo de conocimiento, algo esencial de cara a la atención al paciente.

La necesidad del profesional de mantenerse a la vanguardia en su especialidad médica se hace, por tanto, muy difícil, y no solo por la falta de tiempo, sino también por la dificultad de acceso a la información bibliográfica, la inversión económica que implica consultar revistas científicas o por el hecho de que, en ocasiones carecemos de la habilidad para buscar la mejor evidencia.

Hay un abismo entre lo que el médico pretende o procura leer y lo que la realidad del día a día le permite... "...Sí, he oído hablar del artículo que me comenta, pero no he tenido tiempo de leerlo en profundidad..."

Llegados a este punto y considerando la relevancia de leer revistas científicas como parte de la actividad médica diaria, se me hace pertinente contestar a la pregunta que da título a esta sección.

1.1.1. ¿Por qué y cómo comenzar a leer un artículo científico?

Son diversos los motivos por los que leemos un artículo científico[6]:

- Para estar al día de las noticias profesionales. Muchas revistas sirven como órganos de difusión de asociaciones médicas u otras sociedades científicas.

- Para conocer los avances logrados en una determinada especialidad médica y que se atienden con frecuencia en la práctica asistencial diaria.
- Para saber cómo un investigador de referencia estudia, diagnostica y trata una determinada enfermedad.
- Para conocer en profundidad la robustez, la eficacia, la seguridad y el grado de dificultad de una nueva prueba diagnóstica.
- Para aprender sobre el curso clínico y el factor pronóstico de las comorbilidades, de manera que podría tratarse de algún problema ya definido, cuyo curso y pronóstico va siendo cada vez más claro para el especialista; o bien, podría tratarse de una nueva entidad patológica, para la que conviene estar alerta de las recomendaciones diagnósticas o terapéuticas que se vayan publicando.
- Para conocer nuevos datos sobre la etiopatogenia de la enfermedad y de sus factores de riesgo.
- Para distinguir con criterio sobre las maniobras preventivas, terapéuticas y de rehabilitación que sean realmente eficaces y cuyo coste esté realmente justificado.
- Para entender las nuevas corrientes de discursos, enjuiciamientos y restos que se presentan para cubrir nuevas necesidades en salud, la mejoría de la calidad de la atención sanitaria y la eficiencia de la atención clínica.

Ahora bien, tan importante como saber por qué leemos es planificar bien qué vamos a leer, y los criterios de búsqueda del material, cuyo proceso ha de ser selectivo y riguroso, pero también creativo y abierto.

El saber desde luego sí que ocupa lugar,... ... y lo más importante: nos ocupará tiempo

Debemos plantear de manera lógica y clara la pregunta a la que pretendemos responder y, al mismo tiempo, estar atentos a nuevos interrogantes que pudieran interesarnos según avanzamos en nuestra búsqueda. En una serie publicada y dirigida por el Dr. Domènech Massons sobre la Metodología de la Investigación en Ciencias de la Salud, en la Universidad Autónoma de Barcelona, Fernández y García[7] estratifican la sistemática de la búsqueda de bibliografía en distintas etapas:

- Hemos de formular una pregunta de búsqueda que sea interesante, relevante y original para el propósito que nos ocupa. ¿Cuál es la supervivencia global del melanoma en España? y ¿qué ventajas ofrece la inmunoterapia subcutánea frente al tratamiento sintomático en el paciente con rinitis alérgica? son algunos ejemplos válidos de lo que queremos decir.

- Hemos de seleccionar los recursos –bases de datos– que vamos a utilizar en nuestra búsqueda. En biomedicina, la más utilizada es Medline, a la que se accede libremente vía Internet a través de PubMed (www.pubmed.gov). A continuación, podemos ver los ejemplos de términos de búsqueda en el vocabulario MeSH para los ejemplos sugeridos anteriormente **(tabla 1)**:

Tabla 1. Ejemplos de términos de búsqueda en el vocabulario MeSH

Pregunta de investigación	Términos MeSH
¿Cuál es la supervivencia global del melanoma en España?	*Overall survival* *Melanoma*
¿Qué ventajas ofrece la inmunoterapia subcutánea frente al tratamiento sintomático en el paciente con rinitis alérgica?	*Subcutaneous immunotherapy* *Allergic rhinitis*

- Los repertorios y bases de datos que utilizamos tienen un vocabulario y un lenguaje específico que ha de permitirnos recuperar la información que buscamos de una manera eficaz. Para ello, hemos de traducir la pregunta de la búsqueda al lenguaje documental, utilizando el vocabulario MeSH *(Medical Subject Headings)*: un registro de más

de 22.000 descriptores o términos principales que cubren las distintas áreas y temas de conocimiento en ciencias de la salud.

- La estrategia de búsqueda es el proceso por el que se va diseñando y desarrollando la búsqueda. Una buena búsqueda debe ser lógica, sistemática y reproducible para que nos proporcione el máximo de documentos relevantes existentes sobre el tema que nos ocupe. Los expertos recomiendan que conviene empezar por búsquedas más generales para luego ir utilizando unos términos de búsqueda más específicos[7].

- Cuando hacemos una búsqueda bibliográfica, normalmente accedemos a las referencias de los trabajos recuperados, y en muchos casos, también al resumen, pero para informarnos adecuada y críticamente sobre el tema que nos interesa deberemos recuperar los textos completos de nuestra búsqueda. Algunas revistas permiten un acceso libre a su contenido transcurridos unos años desde su publicación. Para ello, hay diversas bases de datos que pueden manejarse. Este es el caso de Free Medical Journals (www.freemedicaljournals.com), que en la actualidad permite el libre acceso a 4.308 revistas (es el caso de *Allergy*, de libre acceso tras 24 meses de su publicación, y desde 1997).

Nos interesa, por tanto, saber seleccionar los mejores artículos, porque la literatura científica es una de las fuentes en las que se debe basar la práctica clínica, y la mejor práctica clínica es la que mejor aplica la evidencia científica.

Tras definir una sistemática de búsqueda de bibliografía, un aspecto esencial del avance en nuestro *cuerpo de conocimiento* consiste en saber seleccionar aquellos que nos den la mejor evidencia clínica disponible para tomar decisiones sobre el cuidado de cada paciente. Surge aquí, querido lector, un concepto en el que hemos de detenernos si queremos seguir avanzando cuidadosamente sobre los objetivos de este libro: la *medicina basada en la evidencia*.

1.1.2. El proceso de la medicina basada en la evidencia

En noviembre de 1992, de la mano de la epidemiología clínica –entendida por David Sackett et al. como la disciplina para aportar conocimiento científico sobre la práctica clínica–, la *Journal of the American Medical Association* (JAMA) publicaba un manifiesto en el que se decía[8]:

"Está emergiendo un nuevo paradigma para la práctica médica. La medicina basada en la evidencia resta valor a la ex-

periencia clínica no sistemática y al razonamiento fisiopatológico como elementos suficientes para la toma de decisiones clínicas y, en cambio, estimula el análisis de la evidencia proveniente de la investigación clínica."

En realidad, el término "medicina basada en la evidencia" (MBE) había sido publicado por primera vez un año antes por Gordon Guyatt[9], profesor de Epidemiología clínica de la McMaster University en Ontario (Canadá), aunque se utilizó sin ninguna definición explícita. Sin embargo, en 1996, un artículo aparecido en la *British Medical Journal* aclaraba qué era y qué no era medicina basada en la evidencia[10] y, pocos meses después, un libro titulado *Medicina basada en la evidencia. Cómo enseñar y ejercer la MBE*, publicado por Sackett et al.[11] definía y desarrollaba el concepto como:

"... la utilización consciente, explícita y juiciosa de la mejor evidencia científica clínica disponible para tomar decisiones sobre el cuidado de los pacientes individuales".

La MBE ha de ser entendida como un proceso sistemático de búsqueda, evaluación y utilización de los hallazgos de la investigación biomédica como base esencial para la toma de decisiones en la práctica clínica. Es decir, ya no se trata de validar, sino de negar validez a todo lo que no demuestre valía previamente y, para ello, se establece toda una metodología de análisis de la información, que procede fundamentalmente de los ensayos clínicos, pero no de cualquier ensayo

clínico, sino de aquellos realizados de forma prospectiva, controlada y aleatoria.

En los últimos años, la idea de la MBE se ha ido consolidando como un nuevo paradigma, como un cambio en la naturaleza del saber médico y, con ello, un cambio de perspectiva en la práctica clínica y terapéutica. Veamos cómo podemos entender este cambio de perspectiva.

Conforme al antiguo paradigma, los clínicos disponían de una serie de opciones para resolver los problemas clínicos a los que se enfrentaban. Las observaciones no sistemáticas a partir de la experiencia clínica se entendían como una forma válida de desarrollar y mantener los conocimientos acerca del pronóstico del paciente, el valor de los exámenes diagnósticos y la eficacia del tratamiento. El profesional podía reflexionar sobre su propia experiencia clínica, sobre la biología subyacente y los mecanismos fisiopatológicos de la enfermedad, consultar un libro de texto o preguntar a un experto local. Leer las secciones de introducción y discusión de un artículo se consideraba una forma adecuada de obtener información relevante a partir de una revista actual. Este viejo paradigma confería un elevado valor a la autoridad científica tradicional y a la adherencia a los enfoques estándar, basados principalmente en el conocimiento en la materia y en la experiencia clínica.

Sin embargo, los padres de la MBE dicen haber enunciado un nuevo paradigma: un nuevo modelo de cambio que

está permitiendo la transformación de la práctica clínica. El fundamento de este movimiento consiste en que todas las decisiones sobre diagnóstico, pronóstico, etiología y terapia estén basadas en evidencias sólidas, procedentes de la mejor investigación clínica y epidemiológica posible. Al mismo tiempo, un registro sistemático, reproducible y no sesgado de las observaciones aumenta notoriamente la confianza que se puede tener en los conocimientos sobre el pronóstico del paciente, el valor de los exámenes diagnósticos y la eficacia del tratamiento. Conocer bien la base fisiopatológica de la enfermedad está muy bien, pero del todo insuficiente para una buena praxis clínica.

A diferencia del antiguo, el nuevo paradigma confiere un valor mucho menor a la autoridad. Detrás de ello está la creencia de que los médicos, por su habilidad para realizar un análisis independiente de las evidencias, evalúan la credibilidad de las opiniones ofrecidas por los expertos. No obstante, ello no implica un rechazo de lo que se puede aprender de los colegas y de los maestros, que por su experiencia como clínicos gozan de una visión profunda de los métodos para realizar una historia, un examen físico y de las estrategias diagnósticas. Estos conocimientos nunca pueden adquirirse a partir de la investigación científica formal.

En definitiva, practicar la MBE no significa negar las bases del antiguo paradigma, sino incorporar progresivamente a la práctica clínica diaria los postulados del nuevo, recordando siempre que la mejor atención clínica es la que aplica mejor la evidencia científica.

Una última presunción del nuevo paradigma es que los médicos, cuya práctica se basa en la comprensión de las evidencias subyacentes, proveerán una asistencia superior a los pacientes. Por ello, la pregunta que surge es obvia: ¿cuál es la práctica más efectiva y eficiente en una situación concreta? Para cumplir con este nuevo modelo de MBE es necesario comprender algunas normas relativas a la evidencia, para interpretar correctamente las publicaciones acerca de las causas, pronóstico, exámenes diagnósticos y estrategias terapéuticas. De dicho razonamiento derivan las etapas de la práctica clínica basada en la evidencia.

La lectura crítica de un artículo científico es una de las etapas de la medicina basada en la evidencia

La práctica de la MBE se articula en torno a una secuencia de pasos:

- La formulación de preguntas estructuradas susceptibles de ser respondidas. Identificar el problema que nos preocupa y elaborar preguntas susceptibles de respuesta nos permitirá convertir las incertidumbres clínicas en preguntas que puedan ser respondidas mediante la investigación científica.

Para evaluar la relevancia se requieren varios parámetros, los cuales tienen que ser delimitados antes de iniciar la búsqueda bibliográfica. Una pregunta clínica bien estructurada contiene 4 elementos que deben ser bien definidos. Estos elementos pueden resumirse en el acrónimo PICOT, en el cual el quinto elemento es opcional.

P: Problema y paciente de interés. ¿Qué tipo de pacientes o de población? ¿Qué patología o condición de interés?

I: Intervención (o exposición, en las preguntas sobre etiología o factores pronósticos) que estoy considerando como opción para aplicar en mi o mis pacientes.

C: Comparación o intervención alternativa (solo si procede,) en el caso de existir una intervención o abordaje alternativo a ese mismo problema.

O: Objetivo o resultados que quiero considerar para valorar su impacto. Indicador con el que se quiere medir

el impacto de una intervención o la validez de una prueba. ¿Qué esperamos que le suceda al paciente?

T: Tiempo en el que evaluará el impacto de la intervención.

- **Localizar la mejor evidencia de la literatura científica**. Este paso requiere saber buscar con exhaustividad y eficiencia los trabajos científicos desarrollados sobre nuestro tema de interés (búsqueda bibliográfica; véase apartado inmediatamente anterior).

- **Lectura crítica**, que es la parte que da título al libro que nos ocupa. También denominada evaluación crítica (*critical appraisal*) de los documentos obtenidos. La MBE incluye la lectura crítica de artículos científicos entre sus pilares básicos. Evaluar críticamente un artículo científico significa valorar su grado de evidencia, o lo que es lo mismo, saber valorar críticamente la validez –aproximación a la verdad– de dicho artículo, saber interpretar los resultados y poder determinar su utilidad para cada paciente –el potencial de aplicabilidad clínica–.

- **Aplicación de las conclusiones a la práctica clínica**, teniendo en consideración los riesgos y beneficios, las expectativas, la preferencia de los pacientes y sus necesidades emocionales. Los profesionales expertos en un tema concreto intentan aplicar las mejoras mediante protoco-

los de asistencia a partir de la adaptación o desarrollo de las Guías de Práctica Clínica (GPC) basadas en la evidencia científica. Las GPC tienen la potencialidad de reducir la variabilidad y mejorar la práctica clínica.

En España, la puesta en marcha en 2006 del Programa de Elaboración de Guías de Práctica Clínica en el Sistema Nacional de Salud (SNS), coordinado por GuíaSalud (www.guiasalud.es), ha supuesto un salto cualitativo en el desarrollo de GPC en nuestro medio. Este programa, mediante un convenio ministerial entre la Agencia de Calidad del SNS y las agencias y unidades de Evaluación de Tecnologías Sanitarias, se comprometió a la elaboración de una metodología común tanto para la elaboración de GPC como para su implementación y su actualización.

- La última etapa de la MBE consiste en evaluar los resultados obtenidos en la propia práctica y contrastarlos con los de referencia.

La MBE está aportando numerosos beneficios a la terapéutica[12]: *a)* la conciencia cada vez más generalizada sobre el hecho de que las decisiones se sustenten en una base científica sólida; *b)* la MBE responde más adecuadamente a los cambios en el concepto o naturaleza de lo que se entiende por evidencia clínica –la población y no el paciente individual es ahora la "unidad estándar de información clínica analiza-

ble" y la bioestadística y la epidemiología clínica los métodos más adecuados para obtenerla–; *c)* la MBE permite afrontar mejor los retos derivados de la exigencia de la evaluación de la calidad asistencial.

Complementar la evidencia científica con la experiencia clínica del profesional médico y proporcionar al paciente una información adecuada son los tres ejes sobre los que se sujeta la toma de decisiones clínicas. Como tan acertadamente expusiera Santiago Ramón y Cajal hace algo más de un siglo:

"Alejados, el dato experimental y el juicio médico apenas se prestan ayuda; asociados en el mismo intelecto se iluminan y fecundan mutuamente".

1.2. Estructura básica de un artículo científico

A continuación, vamos a identificar la estructura y los elementos que dotan a un artículo científico de coherencia y rigor metodológico.

Day y Gastel definen artículo científico como un informe escrito y publicado que describe resultados originales de investigación[13]. Normalmente, el escrito que se va a publicar es definido por la tradición médica, la práctica editorial, la ética científica y la influencia de los procedimientos de impresión y publicación[14].

La parte preliminar de un artículo científico incluye el título, los autores, el lugar de realización y el resumen

La primera acción que el lector debe considerar es el **título del artículo**, con la intención de apreciar si el mismo sugiere que el contenido será útil o interesante para su propia práctica clínica. De ahí la importancia que tiene para los autores de un trabajo elegir un título que identifique con precisión el tema principal del artículo, y ofrezcan en pocas palabras la idea de lo que el lector hallará en dicho artículo[15]. La estructura del título ha de ser sencilla y cumplir las siguientes características:

- Atractivo, indicativo del problema investigado.
- No muy largo: 15 palabras máximo, redactadas en tono afirmativo. Está demostrado que este es el número de palabras en una frase que marca la capacidad de retención inmediata que tiene el lector medio. Al leer una frase de más de 20 palabras, empezaremos a perder el significado de las primeras palabras que ha leído. Hemos de volver entonces a leer la frase, y esta es la mejor prueba de su longitud excesiva.

En 1952, Robert Gunning creó el índice de niebla para medir la claridad o la dificultad de un texto. Se fundamenta en dos cuestiones básicas: 1) *la longitud de las frases, y* 2) *la longitud de las palabras. Así, el índice de Gunning se mide como*

*IDN = [(nº medio de palabras por frase + % de palabras con 3 sílabas)*0,4]. En esta escala, 16 es un nivel de dificultad que corresponde a textos de lectura difícil, y el nivel 11 es propio de textos legibles para cualquier público. Para que nos hagamos una idea, la Biblia, Shakespeare y Mark Twain tienen un valor de IDN en torno a 6. Para calcular el índice de Gunning de un texto, recomiendo visitar la dirección http://simbon.medpage.com/Fog/.*

- Debe proporcionar información sobre: los sujetos (grupos) que se estudian, las características de los mismos y el ámbito donde se realizó la investigación.

- Único párrafo sin puntuaciones, claro, conciso y preciso sobre los contenidos del artículo.

También hemos de prestar atención a los **autores del artículo**, ya que si son conocidos y su trayectoria corresponde a un trabajo serio, cuidadoso y continuado, es probable que valga la pena leerlo16. No obstante, en la siguiente sección de este capítulo (1.3 Los falsos mitos en la publicación de artículos), desenterraremos algunas de las falsas presunciones que se tienen a la hora de leer un artículo científico de un autor determinado.

Para el Dr. Díaz Portillo, autor de la *Guía práctica de lectura crítica de artículos científicos originales en Ciencias de la Sa-*

lud, los autores han de cumplir una serie de requisitos necesarios y suficientes al iniciar la publicación de cualquier investigación en grupo[17]:

- **Haber hecho una contribución intelectual significativa al proceso de la investigación.**
- **Estar capacitado para presentar y defender públicamente la totalidad del informe (responsabilidad pública del contenido total del artículo).**
- **Haber participado en la redacción o revisión del informe final de investigación.**

Es interesante identificar el **lugar en el que se ha llevado a cabo el trabajo** y valorar si es lo suficientemente parecido al nuestro como para hacer probable que los resultados, si son válidos, puedan ser aplicados a nuestra práctica clínica[18].

En el **resumen de un artículo científico** se debe hacer referencia a las ideas y conceptos más importantes del artículo, a través de una expresión reducida a términos sucintos, precisos y esenciales. Con un estilo claro, conciso, sencillo e impersonal, debe cumplir una extensión de 150-250 palabras. Se trata de la tarjeta de presentación del artículo al lector, cuya finalidad es la de permitirle identificar de una forma rápida y exacta el contenido del documento. Ha de situarnos con agilidad en la finalidad del estudio, cómo se hizo,

los resultados principales y su grado de significación clínica y estadística, así como las conclusiones que se pueden extraer de los resultados. En definitiva, el resumen debe cumplir una estructura determinada:

- Asegurar los fundamentos y objetivos principales del estudio.
- La adecuación de la metodología: diseño o tipo de estudio, contexto (lugar, tiempo, ámbito); la representatividad en cuanto a número y características de los sujetos participantes, y el grado de intervención del investigador (variable de predicción/independiente y de resultado/dependiente).
- Destacar la relevancia de los resultados o hallazgos principales con valores numéricos precisos.
- Aplicabilidad de las conclusiones principales del estudio.

En muchas revistas se solicita un resumen estructurado, en el que los encabezados son: introducción; objetivos; material y métodos; resultados; discusión, y conclusiones.

Estos son los puntos que corresponden con el cuerpo del artículo y son los que vamos a presentarte a continuación, pero atendiendo al modelo general del proceso de investigación **(figura 1)**.

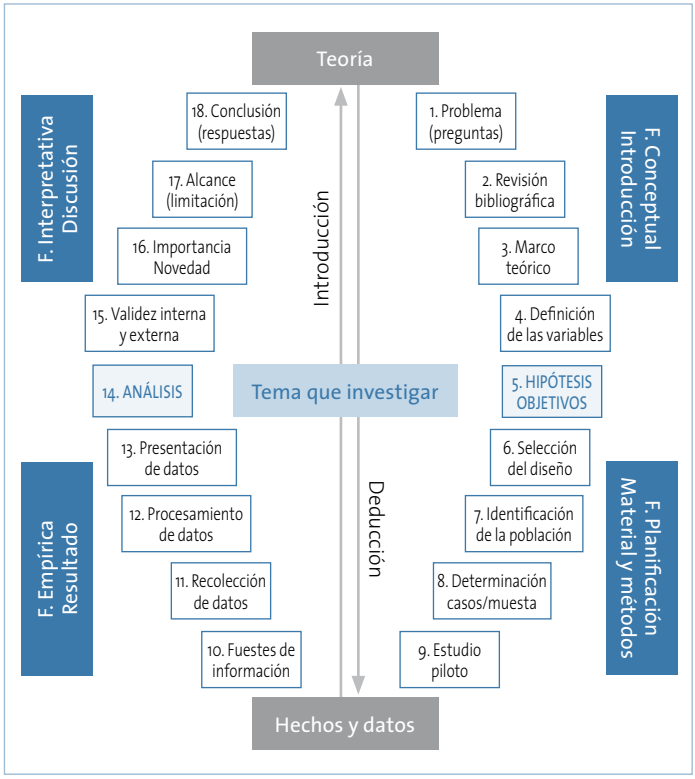


Figura 1. Modelo general del proceso de investigación (modificada de Bobenrieth [2001]).

1.2.1. Parte conceptual: 'Introducción' y 'Objetivos'

La *Introducción* describe los elementos de la fase conceptual de la investigación. Responde a las preguntas de la investigación: el qué y el porqué de la misma. Es donde se define el marco teórico del artículo, por el que se describen los conceptos básicos necesarios para la futura comprensión del artículo original. Es el auténtico soporte conceptual de la investigación, en el que quedan expuestas las ideas procedentes de experiencias y conocimientos anteriores, que son las que proveen los fundamentos del estudio. Se ha de demostrar que se conoce el estado actual del problema, sus antecedentes y los conocimientos más recientes existentes sobre el mismo (revisión bibliográfica). En la *Introducción* se muestra la justificación de las medidas que se van a llevar a cabo a lo largo de la investigación.

La introducción de un artículo en medicina es una vía de transición entre el mundo del lector y el mundo del autor, que despierta el interés del primero por conocer el resto del artículo.

La *Introducción* obedece a la siguiente estructura secuencial:

- Enumeración de los temas generales que abarcan el problema (marco teórico).

- Revisión de los antecedentes del problema. Deben quedar revisadas las referencias históricas y actuales. Este es un concepto amplio que abarca el conocimiento de las novedades existentes en la práctica clínica, nuevas técnicas, nuevos fármacos, etc.

- Definición del problema (pregunta) de investigación. Al leer la introducción de un artículo hemos de ser capaces de discernir si el estudio tiene las características de una buena pregunta de investigación.

Para saber si un estudio tiene las características de una buena pregunta de investigación, podemos utilizar el criterio FINER de Cummings[20].

Factible: valorar los límites de la idea antes de dedicarle tiempo (n.º de sujetos, experiencia técnica de los investigadores, coste en tiempo y dinero, alcance de la investigación).

Interesante: No solo para el investigador implicado.

Novedosa: Se trata de una idea original que aporta información nueva o completa y mejora la ya existente.

Ética: No entraña riesgos para los pacientes, no se precisa de técnicas invasivas, no implica hospitalización ni un mayor número de consultas.

Relevancia: Con consecuencias a nivel social (coste-efectivo), sanitario (modifica la práctica clínica) y del conocimiento (abre la puerta a futuras investigaciones).

- Enunciación y ubicación de las variables (predictora y resultado) que considerar con relación al problema.
- Formulación de los objetivos del estudio.
- Importancia y alcance del estudio.
- Limitaciones del estudio.

Un objetivo es un resultado concreto y previsto que se desea conocer mediante una investigación de tipo descriptivo. **Los objetivos que se enumeran en un artículo científico** definen un estado o situación cuantificable en un lugar y tiempo determinados, a través de un resultado unívoco, preciso, factible y medible que se obtendrá al finalizar el estudio. El objetivo es doble, ya que, por un lado, se facilita la información básica del tema, y por otro, se enuncian las razones para justificar la ejecución del estudio[21].

1.2.2. Parte de diseño: 'Material y métodos'

La redacción de *Material y métodos* de un artículo científico incluye las características del diseño del estudio, describe cómo se va a llevar a cabo el estudio y contiene suficiente cantidad de información como para que otro investigador pueda reproducir el estudio.

La investigación científica busca fundamentalmente el determinar la adecuación o no de ciertas hipótesis a la realidad mediante la observación o la experimentación. Es en *Material y métodos* donde se describe cómo se va a medir el grado de adecuación a la pregunta de investigación concretada en los objetivos; la precisión en el diseño, que aumentará al disminuir la variabilidad inducida por covariables extrañas (factores de confusión); el poder estadístico o capacidad de un diseño para describir las relaciones entre las variables; la ausencia de sesgos, que distorsionan los resultados, y las limitaciones y aspectos éticos del plan de investigación.

La sección de *Material y métodos* de un artículo científico se estructura como sigue:

- Tipo de diseño. Los diferentes tipos de estudios dependen de la clase de pregunta de investigación clínica **(tabla 2)**. Si bien, más adelante pasaremos a describir en detalle la cuestión de los tipos de estudios (véase sección 2.2.1 *Tipos de estudio en investigación clínica*).

- Descripción de los sujetos y cómo son seleccionados: principales características de los pacientes –edad, sexo, procedencia, gravedad, años de evolución de la enfermedad, patologías asociadas, etc.–; número de pacientes incluidos, junto con la enumeración de los criterios de selección (inclusión/exclusión); periodo de seguimiento de los pacientes, terapia previa recibida y terapia concomitante y otras características relevantes.

Tabla 2. Tipos de estudios en investigación clínica según la pregunta de investigación. Modificada de Llorca y Delgado (2009)[22]

Pregunta	Diseño
Eficacia de un nuevo tratamiento	Ensayo clínico
Validez de un método diagnóstico	Estudio transversal Estudio de casos y controles
Validez de un método de cribado	Estudio de casos y controles Estudio de cohortes Ensayo clínico
Etiología de una enfermedad	Estudio de cohorte Estudio de casos y controles
Factores pronósticos de una enfermedad	Estudio de cohortes

- Descripción de la intervención y definición del punto final de la investigación (estudios experimentales). Por ejemplo, si se analiza la eficacia de un fármaco: dosis, vía de administración, duración del tratamiento, posibles interacciones con otros agentes.

- Variables incluidas y cómo se miden los resultados. Definición operacional de las variables independientes (con-

troladas por el investigador) y de la variable dependiente (efecto) y de las covariables incluidas en el estudio (factor de estudio y criterio de evaluación).

- Cómo se han controlado los sesgos. Los grupos que se van a estudiar en cualquier trabajo de investigación han de ser previamente comparables (sin *sesgo de selección*) y han de ser evaluados con las mismas variables, estrategias y técnicas de medición, para evitar lo que se conoce como *sesgo de información*.

- Plan de análisis estadístico: se hace referencia al tamaño de la muestra o *poder del estudio*, pérdidas de seguimiento, interpretación de la "*p*", Test de normalidad –contraste de hipótesis–. Esta es una parte muy importante, ya que sin ella no podrá juzgarse si los hallazgos son respuestas válidas a la hipótesis planteada, o se trata de simples consecuencias del azar[23].

- Limitaciones y consideraciones éticas de la investigación: se ha de describir el procedimiento para obtener el consentimiento informado; la investigación ha debido ser previamente revisada por algún consejo o comité de ética de la institución donde se ha ejecutado el estudio, así como de los posibles riesgos potenciales, si los hubiera, de los sujetos participantes en el estudio.

Como se ha podido apreciar, la sección de 'Material y métodos' es casi la columna vertebral de la publicación de un artículo científico, ya que en ella quedan expuestos todos los elementos que permitirán dar crédito a las conclusiones. A modo de orientación práctica podríamos decir que al finalizar la lectura de los 'Objetivos' y de 'Material y métodos', ¡usted mismo podría ser capaz de enunciar las posibles (no las probables) conclusiones del estudio!

1.2.3. Parte empírica: 'Resultados'

En caso de aprobar los procedimientos de trabajo científico descritos en el apartado anterior, el lector pasará a la sección de *Resultados*. En los *Resultados* de un artículo científico es donde se sustentan las bases para el razonamiento empírico de la decisión clínica. Las decisiones clínicas se fundamentan en un pensamiento científico, con un componente deductivo –basado a partir de los conocimientos teóricos de los mecanismos de las distintas enfermedades y de la actuación de los tratamientos– y de otro empírico, que se basa en los datos (resultados) recogidos de la experiencia clínica derivada del manejo clínico o tratamiento de otros pacientes[24].

En *Resultados* se describe objetivamente lo que ha ocurrido con los pacientes. Debe hacerse una descripción de los resultados, de las diferencias entre grupos de estudio y de la significación estadística y clínica.

La parte empírica de los resultados obedece a la siguiente estructura:

- La primera información, ya sea en formato texto o tabla, debe exponer el tamaño muestral efectivamente alcanzado, seguido de las características demográficas y nosológicas del grupo o grupos de participantes y su comparación.

- Luego deberá presentar la información, preferentemente tabulada, que exprese la relación final observada entre las variables principales con los datos recogidos de todos los pacientes reclutados en origen. De lo contrario, la interpretación de dichos resultados sería arriesgada, y podría llegar a rechazarse su publicación[25].

- Se han de exponer las relaciones exploradas y, si corresponde, los resultados corregidos por variables de control, es decir, aquellas variables que se considera que tienen potencial como para alterarlos.

Los resultados deben estar consignados mediante las medidas de frecuencia, efecto, repercusión (con sus intervalos

de confianza), originalmente seleccionadas. En la sección de *Resultados* se debe presentar también la información obtenida al desarrollar el estudio, en el que, además del texto, se pueden presentar los datos, utilizando formas estilísticas secundarias (tablas, diagramas, gráficos, fotografías, cuadros y figuras). Se observa si los resultados tienen que ver con las cuestiones investigadas y con los propósitos del estudio; si las pruebas estadísticas utilizadas son apropiadas para contestar los interrogantes y si se informa de los valores originales (media, desviación estándar, proporciones) y no solo los resultados de las pruebas estadísticas[26].

De la lectura de los resultados, el lector debe elaborar sus propias conclusiones, intentando contestarse si la dirección hallada de los efectos y su magnitud son de interés para su práctica clínica y qué repercusión pueden tener sobre la misma.

1.2.4. Parte interpretativa: 'Discusión' y 'Conclusiones'

La *Discusión* de un artículo científico responde a la pregunta de ¿cuál es el significado de los hallazgos del estudio? La *Discusión* ha de cubrir tres partes fundamentales: *a)* se resumen, analizan e interpretan los resultados de la investiga-

ción de acuerdo con la metodología empleada; *b)* se discuten los aspectos nuevos y relevantes del estudio, sus implicaciones y limitaciones y se discuten y contrastan las hipótesis, y *c)* se elaboran las conclusiones o respuestas concretas a la pregunta (objetivo) de la investigación.

A través de la *Discusión*, los autores de un artículo científico han de convencer al lector de la fiabilidad y *validez interna* de los resultados y ratificar la ausencia de sesgos que puedan contaminar los resultados y las conclusiones (*validez externa*); han de destacar la importancia *(relevancia)* del estudio, si realmente representa un campo nuevo de aplicación a la práctica clínica del profesional (nuevo método de diagnóstico, una nueva vacuna, etc.) y el grado de utilidad clínica del mismo para la práctica asistencial habitual.

Para Díaz Portillo, la *Discusión* de un trabajo científico se basa en la estructura que sigue a continuación[17]:

- **Comentar los resultados propios.**
- **Corroborar la fiabilidad y validez interna y externa de la investigación.**

Una de las contribuciones más importantes de un experimento es que ayuda a excluir las diferentes variables que podrían explicar un fenómeno. Pero un experimento determinado no excluye necesariamente todas las explicaciones alternativas posibles. Desde un punto de vista metodológico, el mejor experimento es aquel cuyo diseño excluye las explicaciones alternativas de los resultados. Idealmente, los resultados de un experimento tendrán solo una interpretación: la variable independiente (predictora o causa) es la responsable de los cambios en la dependiente (resultado).

Las variables extrañas que puedan afectar a los valores obtenidos en la variable resultado suponen una amenaza para la validez experimental. Un experimento resulta válido en la medida que los resultados pueden ser atribuidos a la variable causa, y en el grado que se generalizan fuera del experimento. Si los resultados de un experimento pueden ser atribuidos inequívocamente a la variable causa, se dice que el experimento es internamente válido. El grado en que permite la generalización se refiere a los aspectos relativos a la validez externa. Es decir, la validez externa nos ayuda a responder cuestiones como: ¿a qué población y en qué medios y condiciones pueden aplicarse los resultados del estudio? ¿Se pueden generalizar o extrapolar los resultados obtenidos de la muestra estudiada a la población de donde proviene?

- Interpretar los hallazgos no esperados.
- Confrontar los datos de la investigación con otros similares.
- Reconocer las limitaciones del estudio y confirmar la ausencia de sesgos (autocrítica del estudio con honestidad).
- Discutir las implicaciones teóricas y/o prácticas del estudio (valorar el alcance de la investigación).
- Sugerir nuevos estudios concretos sobre el problema de investigación. Marcar futuras líneas de investigación, que tratarán de dar continuidad al estudio de nuestra pregunta de investigación.

La *Discusión* suele finalizar con la enumeración de las *Conclusiones*. La conclusión representa la respuesta del investigador, apoyado en la interpretación de los resultados obtenidos, a la pregunta de investigación planteada en forma de objetivo. Es decir, el contenido de las Conclusiones corresponde sin duda al contenido de los objetivos.

La parte final de un artículo científico incluye los 'Agradecimientos' y la lista de 'Referencias' o 'Bibliografía'

En *Agradecimientos* se hace un reconocimiento a la institución o a las personas que han colaborado en el trabajo, ya sea de forma intelectual o material, con el autor en su investigación o en la redacción y/o revisión del manuscrito en aspectos de su contenido, estructura o estilo[17].

Finalmente la lista de *Referencias* o *Bibliografía* muestra cómo los autores han buscado e investigado el trabajo de otros autores sobre el tema, incluye todos los artículos científicos consultados. Cada artículo se cita a lo largo del texto, fundamentalmente cuando se refiere a un resultado, criterio emitido en la discusión o una técnica utilizada. También puede referirse a un artículo para ratificar sus resultados o para cuestionarlos[6].

La bibliografía para la preparación de un artículo científico debe seleccionarse según los criterios de medicina basada en la evidencia y resulta necesario seguir una serie de recomendaciones prácticas.

Conviene, por tanto, seguir las siguientes recomendaciones a la hora de preparar la bibliografía de nuestro manuscrito[27]:

- Las referencias deben enumerarse consecutivamente según el orden de aparición en el texto, en las tablas y en las leyendas de las figuras.
- Se recomienda utilizar números arábigos en superíndice y sin paréntesis.
- Se incluyen solo aquellas referencias consultadas directamente por el investigador, no se deben hacer citas de cita (o segundas fuentes).

- Se deben consultar los documentos completos; debe evitarse citar lo que se conoce como bibliografía gris o no localizable –datos no publicados, informes internos, libros de resúmenes de comunicaciones, tesis doctorales, conferencias no publicadas y artículos y editoriales de periódicos–.

- Los documentos que se citen deben ser actuales, salvo por motivos históricos o si no encontrásemos referencias actualizadas.

- Los títulos de las revistas deben abreviarse según el estilo que utiliza la National Library of Medicine (NLM).

La forma y el orden de construcción de las citas dependen de cada revista, aunque hay una serie de *Requisitos* de *Uniformidad* para manuscritos presentados en revistas biomédicas, más conocida como *Normas de Vancouver*.

Se trata de un conjunto de normas para la presentación de artículos para su publicación en revistas médicas. Es el resultado de una reunión de los principales editores de revistas médicas, celebrada en Vancouver, Canadá, en 1978. Su intención era establecer unas directrices sobre el formato que debían tener los manuscritos enviados a sus revistas. Estas directrices o requisitos se publicaron por primera vez en 1979 con el título de Uniform Requirements for Manuscripts Submitted to Biomedical Journals. En ellos se incluyeron los formatos para

las referencias bibliográficas elaborados por la National Library of Medicine (NLM) de Estados Unidos. Este grupo, que llegó a ser conocido como el Grupo de Vancouver, creció y evolucionó hasta convertirse en el International Committee of Medical Journal Editors (ICMJE), que se reúne anualmente y aún hoy se ocupa de revisar estas directrices y actualizarlas.

Por tanto, una parte de estos Requisitos de Uniformidad se ocupa de la forma de redactar las referencias bibliográficas, tomando como base el estilo de la NLM. A diferencia de otros estilos de referencias bibliográficas, las Normas de Vancouver no incluyen explicaciones para la redacción de las referencias bibliográficas, sino un listado de ejemplos. Esto también es conocido como Estilo Vancouver (la última versión tiene 42 ejemplos).

La forma y el orden para construir las citas dependen de cada revista. La mayoría pone inicialmente a los autores, empezando por los apellidos y después el título del artículo citado, seguido del nombre de la revista con abreviaturas tal como se señalan en el *index medicus*. Se anotan el año, volumen, número y páginas del ejemplar consultado. En todo este procedimiento deben señalarse comas, puntos o puntos y comas de acuerdo a recomendaciones vigentes. Todo lo anterior no es de elevada relevancia para el fondo del trabajo, pero sí para el estilo de la revista.

Todo lo expuesto en este apartado sobre la estructura básica de un artículo científico exige un conocimiento adecuado de

los principios que rigen la metodología general de la investigación clínica (v. figura 1), ya que sin ellos resulta imposible disponer de los elementos de juicio que permitan llevarla a cabo. Con el tiempo, y esto va dirigido al estudiante de ciencias de la salud, o a aquel titulado que acaba de embarcarse en la apasionante tarea de cultivar su cuerpo de conocimiento a partir de la evidencia, adquiriremos suficiente práctica en la ejecución de este modelo para la investigación, lo que no es tan difícil de alcanzar. Ello nos llevará a concluir que no todos los artículos publicados satisfacen los requisitos exigidos.

Leer artículos que no poseen el potencial adecuado como para ofrecer "verdades científicas" consume tiempo y distorsiona los conocimientos. Es momento entonces de desenterrar los falsos mitos en la publicación de artículos.

1.3. Los falsos mitos en la publicación de artículos científicos

El actual protagonismo adquirido por los medios de divulgación científica hace que se estén convirtiendo no solo en el indicador de referencia de la producción investigadora en todo el mundo, sino en el objetivo –en muchos casos– de esa

labor. Publicar un artículo no solo reporta a los autores la posibilidad de dar a conocer sus hallazgos u opiniones, sino que constituye el patrón calificador de su actividad investigadora. Hacerlo en revistas de prestigio y conseguir ser citados por otros autores el mayor número de veces aumentará su reconocimiento investigador y abrirá las puertas a nuevas posibilidades de trabajo o de financiación. Por todo ello, antes de iniciar un proceso de lectura crítica es importante tener en cuenta la existencia de algunos mitos en torno a la publicación científica que son claramente contradictorios con la realidad objetiva.

1.3.1. Mito 1. Si un artículo ha llegado a ser publicado es porque es válido

El primer mito consiste en creer que el hecho de que la comunicación científica –artículo original, artículo especial, original breve, nota clínica, caso clínico, artículo de revisión o revisión conjunta– se publique en una revista de prestigio es garantía de su validez[17]. En principio, las publicaciones científicas de alto nivel poseen diversos mecanismos de control de calidad. La realidad enseña que si bien es cierto que algunas revistas biomédicas y de salud tienen consejos editoriales que utilizan el *sistema de revisión por pares (peer-review)* o de evaluadores externos e internos de la revista para la selección de manuscritos de artículos originales, no es menos cierto que otras tratan el asunto de la revisión de manus-

critos con una ligereza y negligencia que llegan a veces a la irresponsabilidad.

Imagine el siguiente experimento: Usted escribe un trabajo científico falso, basado en datos falsos, obtenidos de experimentos sin metodología científica ni validación alguna, firmado con nombres falsos de investigadores que no existen, asociados a universidades que tampoco existen y envía ese trabajo a centenares de revistas científicas del tipo "open access" –que ponen su contenido en Internet para ser publicado gratis. ¿Qué cree usted que pasaría?

El biólogo y periodista americano John Bohannon hizo exactamente eso y los resultados publicados por la prestigiosa revista *Science*, son aterradores, por lo menos para quienes nos preocupamos por la credibilidad de los estudios "científicos".

Bohannon escribió un trabajo falso sobre las propiedades supuestamente anticancerígenas de una molécula falsa y supuestamente extraída de un liquen, y envió el trabajo a 304 revistas científicas "open access" alrededor del mundo. El trabajo resultaba tan totalmente falso, y obviamente incorrecto, con errores metodológicos y experimentales tan graves que, para el

propio Bohannon, deberían haber resultado obvios para "cualquier revisor con formación de química de secundaria y con capacidad para comprender una planilla básica de datos de Excel". Además, los nombres de los autores que firmaban el trabajo y de sus respectivas instituciones eran todos imaginarios. A pesar de todo, más de la mitad de las revistas involucradas (157) aceptaron el trabajo para su publicación sin necesidad de corrección. Algunas revistas de los grupos editoriales Sage y Elsevier aceptaron el artículo sin más, incluso fue aceptado en revistas cuya temática nada tenía que ver con lo redactado, tal es el caso del Journal of Experimental & Clinical Assisted Reproduction. *Fue un verdadero escándalo científico, pero apenas uno más de los que se ven con cierta frecuencia.*

Está claro entonces que el sistema de revisión científica de las revistas tiene importantes limitaciones[28]. Los revisores científicos forman parte de la élite científica, pero el grado de subjetividad que entrañan sus recomendaciones de publicación o rechazo se refleja en las frecuentes discrepancias entre distintos evaluadores que ha de resolver el director o el comité editorial de la revista. A pesar de su utilidad, a la postre los mecanismos de control de la ciencia y de las revistas científicas no constituyen una prueba definitiva de calidad de un trabajo. En las revistas de primera línea se publican trabajos que de entrada parecen aceptables, cuyo diseño o interpretación no pocas veces resulta erróneo cuando se

examina con cierta perspectiva. Pero, además, los trabajos con claras deficiencias casi siempre acaban publicados en revistas de menor nivel de exigencia, después de haber sido rechazados una o varias veces.

En conjunto, hay una enorme masa de publicaciones dadas por válidas simplemente porque nadie las leyó o las desmintió tras ser publicadas[19]. De hecho, una lectura crítica de los artículos muestra que mucho de lo que se publica no logra convencer de su validez científica (validez interna y/o externa). Surgen serias dudas de validez con dudas en cuanto a concepto y rigor metodológico. A esto hay que añadir los graves problemas de redacción –falta de claridad, de precisión y de orden lógico– de algunos artículos. Estos artículos nunca deberían haber sido publicados en la forma en que se presentaron; otros deberían haber sido rechazados, y otros, haber recibido una aceptación condicional a una futura revisión.

1.3.2. Mito 2. El prestigio profesional o académico de un autor es garantía de la calidad de un artículo

El segundo mito es la creencia de que el hecho de que un autor tenga prestigio en su práctica profesional es garantía de calidad y de validez en sus publicaciones. En los círculos académicos hay una fuerte tendencia a considerar el prestigio profesional o académico como una garantía de validez.

Sin embargo, en repetidas ocasiones, se ha demostrado lo contrario.

La realidad muestra que el tener una sólida experiencia de práctica profesional en atención de salud (medicina, enfermería, odontología, farmacia, veterinaria, etc.) no constituye en manera alguna garantía de rigor metodológico de un estudio.

La medicina en particular y la investigación en general no son ni ciencia ni arte, sino una práctica que se ejerce bajo la influencia de valores culturales, información científica, habilidad conceptual, habilidades técnicas, relaciones humanas y valores éticos. El proceso general de investigación requiere de una preparación especializada que incluye tanto la clara compresión de sus fundamentos como un manejo básico de su metodología.

Por ello, hemos de mantener un escepticismo sano, de modo que cuanto más importante sea un artículo, más es la desconfianza y mayor aún el esfuerzo por analizar críticamente dicho estudio.

1.3.3. Mito 3. El antecedente de la publicación de un excelente trabajo en el pasado respalda la calidad de la producción actual

Mantenemos erróneamente la convicción de que que un autor haya publicado en el pasado otro trabajo de investiga-

ción relevante, de buen nivel conceptual y metodológico va a ser garantía suficiente de validez científica de una nueva publicación.

El cumplimiento de los estándares internacionales de método, estructura y estilo de las normas de publicación y de un nivel mínimo de calidad se debe considerar individualmente para cada estudio, con independencia de los trabajos anteriores. Es posible que existan mayores expectativas respecto a la calidad del trabajo, pero por las razones mencionadas, esta puede cambiar con el tiempo y las circunstancias.

El profesor Fox escribió un estudio falso sobre una teoría absurda que se presentó para su evaluación a cientos de expertos y científicos en tres versiones: una de redacción complicada, incomprensible, cargada de términos difíciles; una segunda con menos términos complicados, y una tercera simple y directa, con lenguaje y sintaxis comprensibles incluso para alumnos de bachillerato. Aquellos que recibieron la complicada opinaron que era un tema muy interesante que merecía una continuidad en la investigación. Los que evaluaron la versión escrita en un lenguaje de un nivel medio de dificultad opinaron que se trataba de un tema interesante, pero no consideraron que tuviese valor alguno. Y los de la versión más sencilla, simplemente contestaron riéndose a carcajadas, considerando el tema y su desarrollo un dislate fenomenal. ¿Cuál podría ser la moraleja de todo esto? No

hay científico alguno que se atreva a reconocer que no sabe sobre un tema que debería ser de su experticia. Muchos parecen tener miedo a enunciar estas dos simples palabras: "No sé".

Quiere esto decir que hay por ahí muchas revistas *científicas* que de científicas no tienen nada. El hecho de que un trabajo haya sido publicado en una revista no significa que sea correcto. La ciencia, como toda otra actividad humana, por desgracia no está libre del fraude y la tergiversación del material que se publica. Entre el fraude científico y la tergiversación de los datos –propios o ajenos– para ajustarlos a las hipótesis planteadas, hay toda una amplia gama de posibilidades[29]. Los procedimientos van desde la simple y llana invención de resultados, al uso fraudulento de información procedente de otros investigadores, el plagio de lo que otros autores han publicado anteriormente, el auto-plagio o duplicidad de los trabajos del propio autor, o el *adorno* o *maquillaje* de los estudios, omitiendo todo aquello que resulte *molesto*.

Las posibilidades de fraude son casi infinitas. En Inglaterra, por ejemplo, ha aumentado de manera notable el registro de transgresiones en la ética científica, desde el plagio y la falsificación de datos hasta fraudes completos, con la inclusión de publicaciones de material redundante, que alcanza cerca del 13% de los trabajos publicados[31]. La proporción de artículos publicados de forma duplicada en el *British Journal of Industrial Medicine*[32] aumentó del 5% en 1998 al 12% en

1990. Se estima que la prevalencia de artículos fraudulentos en revistas internacionales es de alrededor del 0,2% de los artículos publicados y que puede ser mayor en las revistas pequeñas de poco impacto científico, en las que los autores raramente pueden ser descubiertos porque sus artículos o no son leídos o no son citados por la comunidad científica[33]. En 2005, la base de datos PubMed recibió 67 notas de retractación de artículos fraudulentos y se cree, sin duda, que es una cantidad inferior al total de artículos que se publican con plagio o con invención o falsificación de los resultados.

Ahora bien, ¿cuáles son las causas del fraude, del plagio, de la publicación duplicada y de fenómenos similares? Además de por motivos de índole individual, la presión social y financiera a la que se ven sometidos los investigadores por publicar rápido y con buenos resultados, la dificultad de ser sorprendidos en flagrante y el potente mecanismo de autoengaño consistente en hacer lo que se quiere ver parecen ser las causas más aparentes.

La presión que sufren los científicos para publicar el máximo número de artículos ocasiona fenómenos tales como el denominado "mínima unidad publicable" o "publicación salami", la cual permite obtener el máximo posible de artículos (como rodajas de salchichón) de un determinado trabajo científico, además de que un enorme número de publicaciones tienen una utilidad más que cuestionable.

1.3.4. Mito 4. El investigador por la naturaleza objetiva de su trabajo está libre de prejuicios y de sesgos

Tras un análisis de la historia de la ciencia y un recorrido por el pensamiento de algunos filósofos de la ciencia, queda claro que no existe ningún método que pueda preciarse de ser el *método científico*, y que la realidad no puede estar sometida a un único método de estudio. El científico, en su esfuerzo por comprender el aspecto de la realidad que le interesa, está obligado a alejar de su mente los cánones y presupuestos culturales, y debe procurar mantener la libertad de pensamiento y emplear su ingenio y creatividad para elaborar el procedimiento más apropiado para su investigación. Todo hombre de ciencia se convierte en investigador en el instante en que descubre el mejor método para estudiar el

problema que le interesa desde su enfoque particular. Esta conducta natural del científico ha generado una gran cantidad de *métodos* que hoy son un reto para cualquier sistema de clasificación. Todos estos métodos han de estar libres de sesgos y prejuicios en la naturaleza del investigador.

Se debe a Sterling[34], en la década de 1950, hablar de *sesgos de publicación*, definidos como una tendencia sistemática a que los resultados positivos se publiquen más que los negativos (aquellos cuyas diferencias no alcanzan el nivel de significación estadística), y a que una buena parte de la investigación no se publique de manera que pueda ser adecuadamente valorada[35]. Esta definición asume que el error recae en los investigadores y no en los equipos editoriales de las revistas, lo que no se ha comprobado. Una definición más realista es que el sesgo de publicación se produce cuando lo que se publica no representa el total de las investigaciones realizadas acerca de un tema. Si lo que se publica no representa lo que se investiga se está distorsionando todo el proceso de aprendizaje y progreso en un tema.

Se pueden distinguir *tres tipos de sesgos de publicación*: *a)* el de prepublicación, es el más frecuente que se produce cuando los autores deciden no publicar un estudio porque es negativo; b) el de publicación, que introducen los editores al rechazar un estudio por ser negativo, y c) el de pospublicación, resultado de la exclusión de estudios negativos al realizar una revisión de la bibliografía o un metaanálisis[36].

Tanto la publicación de ensayos clínicos como la restricción del acceso a sus resultados o la omisión de datos importantes pueden desembocar en recomendaciones terapéuticas peligrosas, máxime cuando los resultados de los artículos publicados difieren de los que no se publican. Por ejemplo, el metaanálisis realizado sobre la eficacia de varios fármacos inhibidores de la recaptación de serotonina en el tratamiento de la depresión en niños de entre 5 y 18 años sugiere que la mayoría de ellos son eficaces y seguros. La combinación de los resultados de los ensayos publicados y de los no publicados indica que no son eficaces y que sus riesgos superan los beneficios, por lo que se debería desaconsejar su uso. El sesgo de publicación es, indudablemente, una deformación de la realidad que solo puede llamarse fraude científico.

Son varias las explicaciones que podemos dar ante un sesgo de publicación:

- Un primer mecanismo psicológico nos lleva a considerar el escaso valor científico de la demostración de la inexistencia de una asociación o efecto. En principio, no resulta interesante aquel procedimiento que no demuestre ningún efecto. Si el resultado negativo se oculta, o no se hace público, nunca podrá contrarrestar al resultado positivo. La falta de publicación de los resultados derivados de una investigación es una mala conducta científica. Por otra parte, puede darse un sesgo de publicación provocado

por los evaluadores o directores de revistas que rechazan o son poco favorables a publicar resultados negativos sobre temas de investigación sobre los que ya hay datos que indicaron la existencia de un efecto o asociación positiva.

- Otro mecanismo psicológico que puede explicar el sesgo de publicación es la propensión del investigador a sobrevalorar la eficacia o la efectividad de una técnica y a minusvalorar sus efectos negativos y desventajas.

¡Todos tendemos a ver la realidad de manera que nuestra labor parezca útil y positiva, aunque esto puede llevar a errores graves!

El resultado final de la actividad científica solo se logra cuando el autor o los autores ponen al alcance de la comunidad científica sus investigaciones. No obstante, es necesario combatir la tendencia a "publicar por publicar" o a hacer investigaciones cuyo interés es más que dudoso.

Por todo ello, resulta indispensable desarrollar una serie de criterios objetivos para la lectura o evaluación críticas de los artículos, y en ese sentido existe el consenso de que una manera transparente y objetiva de evaluar un artículo debe referirse al proceso de investigación científica, de tal manera que es posible a través del artículo tener una percepción del proceso de investigación que generó los hallazgos (v. figura 1).

1.4. ¿Qué es la lectura crítica de un artículo y qué se persigue?

La formación del profesional sanitario tiene unas raíces clásicas en las que el conocimiento se busca a través del estudio; es un enfoque tradicional en el que el médico va aprendiendo de manera pasiva la información que se le va transmitiendo. Un paso más evolucionado es aquel en el que el individuo adopta una actitud activa y por su propia iniciativa emprende la búsqueda de información y la analiza, utilizando conjuntamente una serie de claves metodológicas y la experiencia personal, de tal manera que realiza una valoración o análisis crítico de lo que lee[37].

La práctica clínica se ha estado realizando a partir de la intuición, la experiencia clínica no sistematizada y el razonamiento fisiopatológico aislado; todas ellas son circunstancias necesarias, pero no suficientes, para una toma de decisiones óptima.

Tal como se ha descrito anteriormente (v. apartado 1.1), la lectura crítica o valoración crítica de los documentos publicados en medicina es la tercera etapa de la medicina basada en la evidencia.

La incorporación permanente de la mejor evidencia disponible procedente de la investigación clínica es la alternativa propuesta para mejorar la toma de decisiones. Esta evidencia puede extraerse individualmente de la lectura crítica de la literatura científica médica, y por ello ha de considerarse la respuesta individual a la oferta informativa.

La lectura crítica consiste en evaluar e interpretar la evidencia aportada por la bibliografía científica, considerando sistemáticamente los resultados que se presentan, su validez y su relevancia para el propio trabajo[38]. En otras palabras, la lectura crítica nos permite adquirir las habilidades necesarias para excluir con la mayor prontitud los artículos científicos de mala calidad y aceptar aquellos otros con la suficiente calidad científica para ayudarnos en nuestra toma de decisiones para el cuidado de los pacientes.

La finalidad última de la lectura crítica es ayudar a los profesionales a tomar decisiones adecuadas a una situación clínica o de gestión concreta. La valoración crítica de la literatura es, por tanto, el principal precedente del intento de incorporar la información científica más válida y fiable en la práctica clínica. En este sentido, el director de la Unidad de Investigación Clínica del Hospital 12 de Octubre (Madrid), el

Dr. Gómez de la Cámara, sostiene lo siguiente: "La utilización óptima de la información científica contrastada y fiable, aquella procedente de la mejor investigación, sería la clave para mejorar la eficacia de la práctica clínica, reduciendo la incertidumbre y aumentando la estima y el aval ético del profesional[39]".

Las revistas de investigación en alergología, cirugía o cualquier otra especialidad de la medicina pueden contener información que sirva como guía para una práctica clínica concreta: la elección de una determinada inmunoterapia, un procedimiento de diagnóstico o una nueva técnica quirúrgica. Sin embargo, gran parte de toda esta bibliografía es muy extensa y está mal organizada. Hay quien piensa incluso que el lugar más idóneo para guardar el 90% de lo que se escribe en la literatura médica es la papelera "...aun cuando se trate de artículos escritos por profesionales famosos en revistas de prestigio[40]". Aunque no estemos muy de acuerdo con esta afirmación de Altman, sobre todo por tratarse de una reflexión excesivamente dramática y generalista, de lo que no cabe duda es de que un amplio porcentaje de los trabajos que se publican están aquejados de errores de método graves.

Nos quedamos, por tanto, con lo manifestado por Sacket (2001)[41] sobre la calidad variable de los estudios y los documentos de investigación; algunos artículos contienen errores importantes que disminuyen o incluso invalidan las

afirmaciones de los autores, aunque otros errores menores pueden ser inevitables y pueden no disminuir el valor de una investigación rigurosa.

A consecuencia de la calidad variable y de la gran producción de publicaciones científicas, los lectores necesitan cultivar su capacidad para decidir los defectos que debilitan las afirmaciones de los investigadores acerca de los resultados obtenidos. Sin una evaluación crítica de lo que se lee, el lector se encontrará relativamente desamparado a la hora de decidir cuál es la nueva información que deben incorporar a su práctica.

La lectura crítica ha de reducir la variabilidad impuesta por la práctica clínica

El intento de disminuir la variabilidad de práctica clínica, es otra de las motivaciones para ejecutar sistemáticamente la lectura crítica. El sistema sanitario, como sistema de producción, se basa en la toma de millones de pequeñas decisiones, fundamentalmente decisiones que un médico toma delante de un paciente y que implican hacer o no hacer algo. Del agregado de estas decisiones derivan tanto los resultados de salud (a veces adversos) como el gasto sanitario[42]. Una aproximación metodológica posible para soslayar, al menos en parte, estas limitaciones es el análisis de las variaciones

geográficas en la práctica médica. Estos análisis se centran en el lugar donde reside el asegurado, analizan su exposición a los servicios sanitarios y determinan qué parte de la misma no se debe a diferencias en necesidad[43]. En cierto sentido, las poblaciones son cautivas de la cantidad y calidad de los servicios que se prestan en su área geográfica. "En el mercado de la atención sanitaria, la geografía determina el destino: los servicios que se reciben dependerán, en gran medida, de la oferta de servicios disponibles en el lugar donde uno vive y del estilo de práctica de los médicos locales", escribió Wennberg, precursor del análisis de la variabilidad geográfica[44]. La variabilidad en la práctica clínica plantea importantes dudas sobre la indicación de determinadas intervenciones y sobre los resultados para los pacientes. En situaciones de incertidumbre habría que cuestionarse la legitimidad de la aplicación de las valoraciones subjetivas de los profesionales a despecho de la evidencia disponible o las preferencias de los pacientes[45].

La variabilidad en la práctica clínica tiene su origen en tres tipos de factores: dependientes de la población, dependientes de la oferta sanitaria de los servicios del hospital y dependientes del médico. Cada profesional adopta un estilo de práctica médica basándose en sus preferencias, creencias o percepciones sobre el valor marginal de una intervención (efectividad subjetiva). Esto es lo que se denomina Hipótesis de la Incertidumbre Profesional. Todo ello va a traer consecuencias importantes

en los resultados clínicos, en la calidad de vida de los pacientes y en el coste-efectividad del sistema sanitario.

La variabilidad en la práctica clínica va repercutir de forma negativa en la calidad asistencial, poniendo de manifiesto tres tipos de problemas: un problema de efectividad (los resultados pueden variar dependiendo de quién, cómo y dónde se trate el paciente); un problema de eficiencia (manejo indebido de los recursos, con el aumento de coste-oportunidad que esto comporta), y un problema de accesibilidad (criterios variables de diagnóstico o tratamiento). En la reciente, pero ya amplia, literatura sobre la variabilidad en la práctica médica se ha comprobado que hasta el 30% del coste total sanitario de Estados Unidos podría evitarse, sin afectar negativamente a los resultados en salud[46].

En contraposición a todo lo expuesto, hemos de identificar las fortalezas y debilidades de las publicaciones científicas. La lectura crítica es una evaluación objetiva y crítica de las fortalezas y las debilidades de la metodología de una investigación completa publicada. Al conducir una lectura crítica deben valorarse tanto los aspectos negativos como los positivos del informe. Dado que todos los estudios de investigación presentan debilidades y limitaciones, la clave de una evaluación crítica no es solo citar las debilidades específicas, sino, también, evaluar su impacto en la totalidad del estudio.

La lectura crítica de un informe de investigación incluye la evaluación de los aspectos del estudio para juzgar sus méritos, significado, alcance y limitaciones[19].

Identificar la potencia y las deficiencias mediante la lectura crítica de un artículo científico nos va a permitir encontrar las publicaciones de buena calidad, reconocer las publicaciones fiables que apoyan el conocimiento adquirido por el aprendizaje y el pensamiento racional, tomar decisiones clínicas basadas en la evidencia y examinar el valor y la pertinencia de los resultados.

Ahora bien, para llevar a cabo una evaluación crítica de un artículo científico, el lector ha de tener en cuenta que este proceso lleva implícita una serie de requisitos[19]:

- Conocer todas las etapas del proceso de investigación (v. figura 1): conceptual (teórica), planificación (diseño), ejecución (resultados) e interpretativa (significado).

- Tener presente la correspondencia de cada etapa de la investigación con las partes del artículo científico original: conceptual (introducción), planificación (materiales y métodos), empírica (resultados) e interpretativa (discusión y conclusión).

- Tener un dominio básico de la terminología de uso frecuente en investigación y publicación científica.
- Reconocer los requisitos fundamentales del estilo científico: claridad, concisión, precisión, orden/secuencia y sencillez.
- Disponer de pautas de lectura crítica para seguir en cada capítulo del artículo original, adaptado al tipo de investigación (paradigma epistemológico/ metodológico) y su correspondiente diseño.

Las pautas para la lectura crítica de un informe de investigación o artículo original publicado son componentes clave que considerar para la evaluación de una investigación. Una pauta sirve como norma o modelo para realizar la lectura crítica. Todas estas pautas tienen como objetivo común proporcionar una lista de componentes críticos del proceso de investigación y de la estructura y el estilo del trabajo que evaluar.

Una buena pauta va a ser el referente, estándar o norma sobre el "debe ser" del proceso, de su escritura y de su publicación.

Basándonos en su estructura básica (v. apartado 1.2), cualquier artículo científico ha de responder así a una serie de preguntas directrices para la evaluación preliminar de su calidad **(tabla 3)**[47]:

Tabla 3. Preguntas directrices para iniciar la lectura de un artículo

Preguntas (evaluación preliminar del estudio)		
¿Tiene interés el estudio?	Título/Resumen	¿Es el estudio relevante para nuestra búsqueda? ¿Son interesantes los resultados?
¿Por qué se llevó a cabo el estudio?	Introducción	¿Se especifican y justifican los objetivos? ¿Se plantea una hipótesis de partida?
¿Cómo se llevó a cabo el estudio?	Métodos	¿Se detalla cómo se lleva a cabo? ¿Se describe la muestra y el reclutamiento?
¿Qué ha encontrado el estudio?	Resultados	¿Sirve el texto como guía rápida del contenido más relevante? ¿Complementan las tablas y las figuras el texto?
¿Cuál es la relevancia del estudio?	Resumen/ Discusión	¿Se pueden generalizar los resultados? ¿Qué novedad aporta a nuestra práctica clínica?

Tabla 3. Preguntas directrices para iniciar la lectura de un artículo *(continuación)*		
Preguntas (evaluación preliminar del estudio)		
¿Qué interés añadido aporta el estudio?	Introducción/ Discusión	¿Se citan otras referencias de interés? ¿Se plantean otras ideas novedosas?

No obstante, y nos referiremos nuevamente al Dr. Díaz Portillo[17], hemos de valorar otro tipo de preguntas más críticas, asociadas al rigor científico de la metodología de la investigación seguida en el estudio, que nos van a ayudar a ser más críticos con su lectura. Estas preguntas son las que figuran en la **tabla 4**:

Tabla 4. Preguntas para ser más críticos
¿Cuál es la importancia o relevancia de la investigación?
¿Estamos de acuerdo con la forma en que el autor abordó la investigación?
¿Cuán apropiados son los procedimientos y técnicas de investigación utilizados por el autor?
¿Hasta qué punto la evidencia presentada fundamenta los argumentos del autor?

(Continúa)

Tabla 4. Preguntas para ser más críticos *(continuación)*
¿Se presentó la suficiente evidencia como para juzgar los resultados de la investigación por uno mismo?
Los hallazgos presentados, ¿se pueden aplicar a otras observaciones propias o de otros autores?
¿Se pudo detectar algunos patrones o contradicciones en los datos que el autor no mencionó?
¿Las conclusiones representan adecuadamente a la población bajo estudio?
¿Estás de acuerdo con las conclusiones?
¿Las conclusiones generalizan demasiado o son lo suficientemente cuidadosas?
Las interpretaciones y reflexiones utilizadas en el artículo, ¿parecen responder a algún interés particular de los autores?
A tu mejor entender, ¿los datos presentados son sólidos y reconocidos por los pares?
¿Logra el artículo cumplir con los objetivos que planteó en un principio?
Revisa alguna bibliografía sobre el tema del artículo. ¿Consideras que el autor ha citado los trabajos más importantes sobre el tema?

Tabla 4. Preguntas para ser más críticos *(continuación)*
¿Cómo afecta este artículo a las impresiones que tenías antes sobre el tema?
¿Consideras que el artículo está claro y bien escrito?

1.5. Fases de la lectura crítica

La lectura crítica de un artículo científico consta de tres fases que, por un lado, nos permiten comprobar distintos niveles de análisis de un trabajo y, por otro, nos ayudan a emplear mejor el tiempo, facilitando la tarea de eliminar aquellos artículos que no cumplen con la calidad científica deseada.

La primera fase pasa por realizar una selección de los artículos localizados previamente. Mediante la evaluación de ciertos parámetros, podremos eliminar ciertos artículos, y así focalizar nuestro análisis en aquellos que pasen a la segunda fase. Los aspectos que hay que tener en cuenta para seleccionar los artículos de interés son: el título, los autores y el resumen. La selección de los artículos debe responder a una serie de preguntas clave, siguiendo el árbol de decisión esquematizado en la **figura 2**.

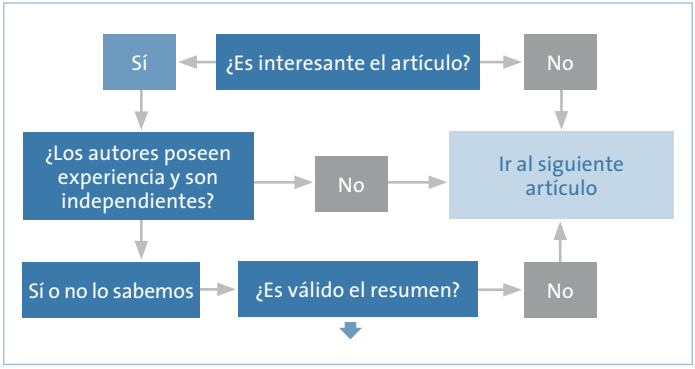


Figura 2. Fase inicial de la lectura crítica: la selección preliminar de artículos (modificada de Díaz Portillo en *Guía práctica de lectura crítica de artículos científicos originales en Ciencias de la Salud*[17]).

La evaluación detallada del título del artículo permite conocer la gran mayoría de las veces el diseño del estudio y sus conclusiones. No es fácil dar un título a un artículo y es, por tanto, una preocupación de los autores elegir un título que efectivamente represente su pensamiento y el diseño del artículo. Especial atención debemos poner porque, como veremos más adelante, muchas veces las conclusiones de los autores no son sustentadas por los datos o el diseño, y esto con frecuencia se reflejará en un título inapropiado.

Los autores deberán aportar credibilidad y experiencia en el tema o, por el contrario, si hay algún motivo para desconfiar de ellos. La observación cuidadosa de los autores nos infor-

mará de si el estudio es la consecuencia lógica de un estudio previo y de la credibilidad de los datos. Recordemos que es imprescindible que la pregunta que el artículo intenta responder tenga sentido biológico, y ello muchas veces resulta de la investigación previa de los autores.

Se debe buscar si el resumen es correcto y, por último, la introducción. En ella, los autores hacen una revisión de la información respecto del tema que se investiga e informan de los objetivos del trabajo. La lectura de la introducción es especialmente útil para el lector, que se ha de plantear una serie de cuestiones básicas sobre si la pregunta de investigación se presenta y justifica adecuadamente. Hay que poner atención para descubrir si se ha omitido información importante, ya que muchas veces los autores ocultan información para hacer que sus datos parezcan más novedosos o evitar antecedentes que estén en contra de sus resultados.

Con esta primera fase, podemos descartar artículos que no se centren en el tema que desea analizar, aquellos que resulten sospechosos sobre su fiabilidad y la de los autores, cuyo resumen no deje claro el contenido del artículo.

En la segunda fase se analizará la capacidad del artículo para contestar al problema que nos planteamos, es decir, se trata de evaluar la validez y utilidad de los resultados. Me-

diante ciertas preguntas podremos descubrir si el artículo se estructura adecuadamente y no quedan elementos de la investigación sin la contestación, aclaración y resolución adecuadas **(figura 3)**.

Los resultados son la parte más importante del trabajo. Habitualmente, los datos más relevantes aparecen presentados en tablas o gráficos. La exposición de los resultados ha de ser lo más clara y objetiva posible, tanto en su redacción como en su presentación. Estos deben ser observados con cautela para ver los datos crudos. Aquí exigiremos o preguntaremos si los resultados están claramente presentados y son rápidamente inteligibles, verificaremos la consistencia

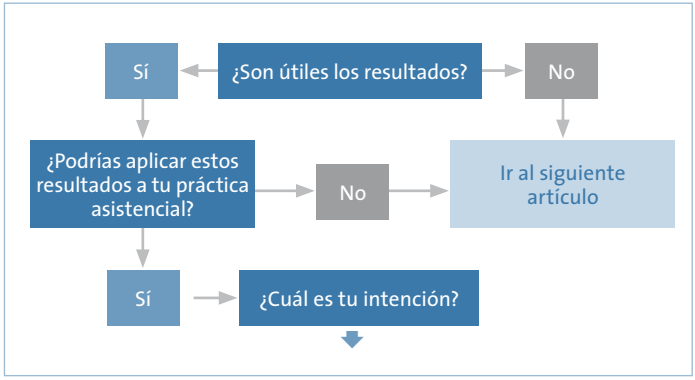


Figura 3. Segunda fase de la lectura crítica: ¿son útiles los resultados? (modificada de Díaz Portillo en *Guía práctica de lectura crítica de artículos científicos originales en Ciencias de la Salud*[17]).

de los mismos, la exactitud y pulcritud de los gráficos y tablas, así como la comparabilidad entre los distintos grupos o poblaciones involucrados en el estudio. Al final de los resultados el lector debe establecer sus propias conclusiones.

En la tercera fase de la lectura crítica de un artículo científico analizaremos en profundidad el material y la metodología utilizados en aquellos artículos que han superado el esquema de trabajo que seguir en las dos fases anteriores.

Llegados a esta etapa de la selección, es pertinente desarrollar una escala cualitativa sobre aspectos metodológicos clave, de manera que exigiremos conocer el grado en que la metodología utilizada se corresponde con la naturaleza de la pregunta de investigación por resolver **(figura 4)**. Partimos, por tanto, de una pregunta clave: *¿Cuál es tu intención a la hora de seleccionar un artículo?* Si lo que pretendemos es evaluar un artículo sobre el uso de una nueva prueba diagnóstica, la pregunta clave deberá ir encaminada a averiguar si hubo una comparación ciega independiente de dicha prueba con un *patrón oro (gold standard, o patrón de referencia)* del diagnóstico. Si lo que pretendemos es aprender sobre el curso clínico y el pronóstico de una enfermedad, la pregunta metodológica clave sería verificar la existencia de una cohorte de individuos correctamente diseñada. Si lo que queremos es saber más sobre la etiología o causalidad de una circunstancia o situación, la pregunta metodológica básica es que exista una demostración sólida de la relación

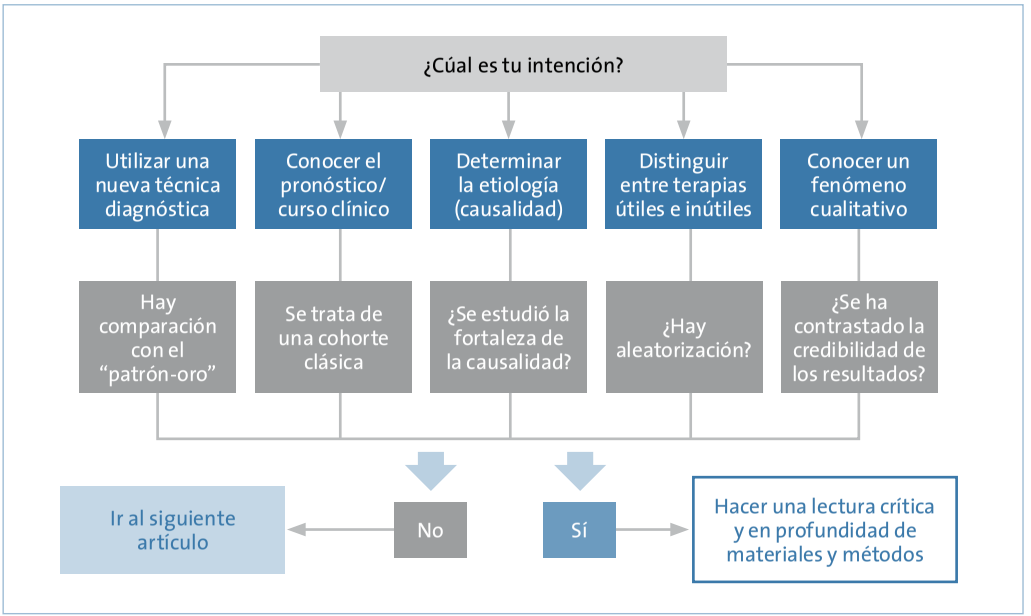


Figura 4. Tercera fase de la lectura crítica: ¿cuál es tu intención? (modificada de Díaz Portillo en *Guía práctica de lectura crítica de artículos científicos originales en Ciencias de la Salud*[17]).

causal en el estudio revisado. Si de lo que se trata es de valorar la validez y utilidad de una intervención o tratamiento, la pregunta metodológica clave consistiría en verificar la existencia de un ensayo clínico correctamente diseñado y con asignación aleatoria de los grupos de estudio.

Una vez que se ha realizado esta selección preliminar, podemos hacer el resto de la revisión clínica desde diversos enfoques:

- Hay aproximaciones objetivas que elaboran escalas cuantitativas o baremos de verificación de la calidad metodológica del estudio en las que se encuentra un número fijo de ítems de aspectos metodológicos que se puntúan según un peso determinado; se suman los puntos de cada ítem para obtener un valor final que pondera la calidad del trabajo.

- Otra alternativa es el uso de escalas cualitativas sobre aspectos concretos de la metodología de un trabajo.

- Y, finalmente, se puede llevar a cabo un análisis de errores o defectos presentes en el artículo. Dichos errores son *mayores* cuando generan un sesgo potencial suficientemente importante como para cuestionar la validez de un resultado (p. ej., no distribuir con procedimientos verdaderamente aleatorios a los sujetos en un ensayo o experimento clínico). Se considera *error menor* cualquier des-

viación de un procedimiento correcto que no amenaza la validez del estudio (p. ej., no describir en qué consiste el tratamiento placebo).

Estas tres fases que acabamos de describir nos permitirán estructurar la lectura crítica y, por tanto, facilitar la evaluación de la calidad del diseño y el contenido de los artículos que nos encontremos y un mejor aprovechamiento del tiempo dedicado a la lectura de los artículos científicos que nos ayuden a resolver el problema planteado en cada momento.

1.6. Niveles de evidencia

A lo largo de este primer capítulo ya hemos hablado de que a la hora de decidir qué artículos son los más convenientes para responder a la pregunta de investigación que nos planteamos, se necesita valorar el nivel de evidencia de dicho artículo.

Tras el primer cribado, son de gran utilidad las escalas de clasificación de la evidencia científica. En función del rigor científico del diseño de los estudios, pueden construirse escalas de clasificación jerárquica de la evidencia, a partir de las cuales se establecen recomendaciones respecto a la adopción

de un determinado procedimiento médico o intervención sanitaria[50]. Aunque hay diferentes escalas de gradación de la calidad de la evidencia científica, todas ellas son muy similares entre sí.

La primera de ellas fue formulada en 1979 por la Canadian Task Force on the Periodic Health Examination para la evaluación de medidas preventivas[51] y adaptada en 1984 por la U.S. Preventive Services Task Force. En su tercera edición, publicada recientemente[52], se evalúa la calidad de la evidencia de una forma más elaborada que no solo tiene en cuenta el tipo de diseño de los estudios **(tabla 5)**. Los grados de recomendación se establecen a partir de la calidad de la evidencia y del beneficio neto (beneficios menos perjuicios) de la medida evaluada **(tabla 6)**.

Ahora bien, estas clasificaciones no están exentas de limitaciones, ya que no pueden distinguir las diferencias de calidad y rigor científico para dos estudios situados en el mismo nivel de evidencia cuando ofrecen resultados opuestos. Para eliminar esta limitación, en nuestro país hay que destacar el esquema de gradación propuesto por la Agència d'Avaluació de Tecnologia Mèdica (AATM) de la Generalitat de Catalunya[53]. Esta clasificación tiene en cuenta, además del diseño de los estudios, una valoración específica de su calidad (**tabla 7**).

Tabla 5. Jerarquía de los estudios por el tipo de diseño (clasificación revisada del Canadian Task Force on the Periodic Health Examination)

Nivel de evidencia	Tipo de estudio
I	Al menos un ensayo clínico controlado y aleatorizado diseñado de forma apropiada
II-1	Ensayos clínicos controlados bien diseñados, pero no aleatorizados
II-2	Estudios de cohortes o de casos y controles bien diseñados, preferentemente multicéntricos
II-3	Múltiples series comparadas en el tiempo, con o sin intervención, y resultados sorprendentes en experiencias no controladas
III	Opiniones basadas en experiencias clínicas, estudios descriptivos, observaciones clínicas o informes de comités de expertos

Tabla 6. Significado de los grados de recomendación (clasificación revisada del Canadian Task Force on the Periodic Health Examination)

Nivel de reconocimiento	Significado
A	Extremadamente recomendable (buena evidencia de que la medida es eficaz y los beneficios superan ampliamente los perjuicios)
B	Recomendable (al menos moderada evidencia de que la medida es eficaz y los beneficios superan los perjuicios)
C	Ni recomendable ni desaconsejable (al menos moderada evidencia de que la medida es eficaz, pero los beneficios son muy similares a los perjuicios y no puede justificarse una recomendación general)
D	Desaconsejable (al menos moderada evidencia de que la medida es ineficaz o de que los perjuicios superan los beneficios)
I	Evidencia insuficiente, de mala calidad o contradictoria, y el balance entre beneficios y perjuicios no puede ser determinado

Tabla 7. Niveles de calidad de la evidencia científica (clasificación de la AATM)

Nivel	Fuerza de la evidencia	Tipo de diseño	Condiciones de rigurosidad científica
I	Adecuada	Metaanálisis de ECA	Análisis de datos individuales de los pacientes Sin heterogeneidad Diferentes técnicas de análisis Metarregresión Megaanálisis Calidad de los estudios
II	Adecuada	ECA de muestra grande	Evaluación del poder estadístico Multicéntrico Calidad del estudio
III	De buena a regular	ECA de muestra pequeña	Evaluación del poder estadístico Calidad del estudio
IV	De buena a regular	Ensayo prospectivo controlado no aleatorizado	Controles coincidentes en el tiempo Multicéntrico Calidad del estudio

V	Regular	Ensayo retrospectivo controlado no aleatorizado	Controles históricos Calidad del estudio
VI	Regular	Estudios de cohorte	Multicéntrico Apareamiento Calidad del estudio
VII	Regular	Estudios de casos y controles	Multicéntrico Calidad del estudio
VIII	Pobre	Series clínicas no controladas Estudios descriptivos: Vigilancia epidemiológica Encuestas Registros Bases de datos Comités de expertos Conferencias de consenso	Multicéntrico
IX	Pobre	Anécdotas o casos únicos	

(ECA: ensayo controlado aleatorizado)

Tabla 8. Nivel de evidencia en función del tipo de estudio (modificada de Abad-Corpa [2002], *Lectura crítica de la literatura científica*[49])

Tipo de estudio	I	II	III	IV	V
Tratamiento	ECA Metaanálisis	Estudios de cohortes o ECA de baja calidad (seguimiento < 80%)	Estudios de casos y controles	Series de casos, estudios de cohortes o de casos y controles de baja calidad (seguimiento < 80%)	Opinión de expertos
Etiología	ECA	Estudios de cohortes o ECA de baja calidad (seguimiento < 80%)	Estudios de casos y controles	Series de casos, estudios de cohortes o de casos y controles de baja calidad (seguimiento < 80%)	Opinión de expertos

Pronóstico	Estudios prospectivos de cohortes con pacientes al inicio de la enfermedad	Estudios retrospectivos de cohortes		Series de casos, estudios de cohortes o de casos y controles de baja calidad (seguimiento < 80%)	Opinión de expertos
Diagnóstico	Estudios de diagnóstico de buena calidad: comparación de la prueba con el estándar de referencia sobre una gran variedad de pacientes (sanos y en estados intermedios de la enfermedad	Estudios de diagnóstico de menor calidad: comparación de la prueba con el estándar de referencia sobre una variedad restringida de pacientes	Comparación independiente y ciega en una variedad apropiada de pacientes, pero el estándar de referencia no se aplica a todos los pacientes	El estándar de referencia no se aplica de forma independiente o ciega	Opinión de expertos

No todos los problemas que se nos plantean en la práctica clínica encontrarán su repuesta en el ensayo clínico aleatorizado o en el metaanálisis (véase el apartado 3.1, *Tipos de estudios en investigación clínica*). Así, por ejemplo, si la pregunta que se plantea es relativa a las preferencias de los pacientes, un estudio observacional bien diseñado será la mejor opción. En cambio, si la cuestión trata sobre la eficiencia, a partir de un estudio riguroso sobre rentabilidad se obtendrá la mejor evidencia[54]. Sin embargo, cuando nuestras preguntas se refieran a la efectividad de una intervención se deben evitar los estudios no aleatorizados[50].

En la **tabla 8** se presenta la forma de decidir qué estudios se deben elegir, analizando su nivel de evidencia, para hacer después una lectura crítica en profundidad y poder responder con mayor acierto a cada tipo de problema[49].

La revisión de la literatura científica debe ir unida a la capacidad de analizarla de forma crítica. Se trata de valorar la evidencia que ofrece la investigación con el fin de poder trasladarla a nuestra práctica asistencial. Para ello, podemos disponer de una serie de instrumentos, que es exactamente lo que vamos a ver a continuación.

1.7. Instrumentos para la lectura crítica

Hablar de los instrumentos disponibles para la lectura crítica de un artículo científico supone profundizar en las tres fases de la lectura crítica. Resulta oportuno entonces que volvamos a hablar del concepto de medicina basada en la evidencia.

Gracias a un grupo de internistas y epidemiólogos canadienses de la McMaster University de Ontario que constituyeron el Evidence-Based Medicine Working Group (EBMWG), en 1992 se publicaron unas guías para usuarios de la literatura científica[8] (v. apartado 1.1), cuya pretensión no fue otra que la de disponer de una herramienta útil que sirviera de ayuda a los clínicos a la hora de decidir cuál es la nueva información que deben incorporar a su práctica cotidiana.

A partir de las pautas de valoración crítica de artículos científicos establecidas por el EBMWG, el National Health Service inglés creó el CASP (Critical Appraisal Skills Programme o Programa de Habilidades en Lectura Crítica). CASP trabaja en colaboración con el Center for Evidence-Based Medicine de la Universidad de Oxford y tiene representación en países como Noruega, Alemania, Países Bajos, Canadá y España. En España existe el grupo CASPe (Programa de habilidades en la lectura crítica español; disponible en http://www.redcaspe.org/) que forma parte de una organización llamada CASP internacional (CASPi), con la que comparte la filosofía docente,

los materiales desarrollados en conjunto, las experiencias docentes y de organización, así como los proyectos de docencia. Al igual que en la versión original inglesa, CASP España tiene la misión de enseñar y difundir habilidades de búsqueda y lectura crítica de la evidencia entre las personas que toman decisiones en los diferentes niveles del sistema de salud (clínicos, gestores, pacientes y ciudadanos) **(figura 5)**.

La organización CASPe se reparte a través de un centro de coordinación en Alicante y otros localizados en distintos puntos de España. Esta estructura permite compartir habilidades y tareas, así como hacer las lecturas más próximas sobre las necesidades y circunstancias locales.

Las guías para el usuario de la literatura científica propuestas por el EBMWG fueron refundidas posteriormente por

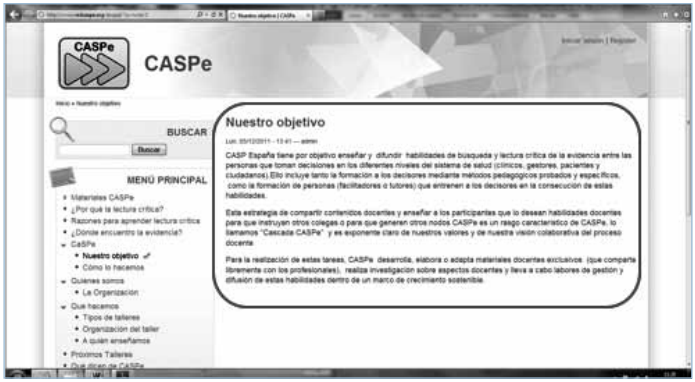


Figura 5. Página de objetivos en http://www.redcaspe.org/.

CASP y se han propuesto parrillas sencillas de preguntas para hacer más accesible la lectura crítica de diferentes tipos de estudios. El logotipo de CASP (v. figura 5) son tres flechas consecutivas que representan los tres pasos necesarios para usar la evidencia en la práctica profesional: **buscar** la evidencia, **valorar** la evidencia y **aplicar** la evidencia.

Lógicamente, se buscó la uniformidad a la hora de presentar dichas parrillas, organizando el cuestionario en torno a tres preguntas básicas: *a)* la evaluación de la validez interna del estudio: *¿son válidos los resultados del estudio?*; *b)* la relevancia de los resultados: *¿cuáles son los resultados?*, y *c)* la utilidad práctica, ya que no solo debemos valorar la proximidad a la verdad de los resultados del estudio, sino también su importancia clínica y su factibilidad: *¿me serán útiles los resultados para atender a mis pacientes?*[55].

En la **tabla 9** se exponen algunos de los listados de preguntas directrices y otras más críticas correspondientes a los diferentes tipos de estudios en los que se ha llevado a cabo una investigación cuantitativa (investigación de fenómenos que se prestan para la medición objetiva y cuantificación precisa; p. ej., edad, IMC, glucemia, colesterol sérico). Desde luego, lo ideal sería poder elaborar respuestas firmes con afirmaciones o negaciones rotundas, pero, por desgracia, esto no suele ser así. Hemos de aprender, por tanto, a interpretar esa zona oscura de la que siempre resultará más difícil tomar las decisiones adecuadas, teniendo en cuenta el grado de cer-

Tabla 9. Valoración crítica de estudios cuantitativos CASP

Ensayo clínico	Prueba diagnóstica
¿Son válidos los resultados del ensayo?	
¿Se orienta el ensayo sobre una pregunta claramente definida?	¿Existió una prueba de referencia adecuada?
¿Se realizó la asignación de los pacientes a los tratamientos de manera aleatoria?	¿Incluyó la muestra de pacientes un espectro adecuado de paciente?
¿Fueron adecuadamente considerados hasta el final del estudio todos los pacientes que entraron en él?	¿Existe una adecuada descripción de la prueba?
¿Se ha mantenido un diseño "ciego" respecto al tratamiento, tanto de los pacientes, clínicos y personal del estudio?	¿Hubo evaluación "ciega" de los resultados?
¿Eran similares los grupos al inicio del ensayo?	¿Influyeron los resultados de la prueba objeto de evaluación en la realización del estándar de referencia?

Etiología	Pronóstico
¿Son válidos los resultados del ensayo?	
¿Se han utilizado unos grupos de comparación claramente identificados que sean similares en cuanto a los factores determinantes importantes del resultado a parte del que se investiga?	¿Fue una muestra representativa y bien definida de pacientes en un momento similar en el curso de la enfermedad?
¿Se han evaluado las exposiciones y los resultados de la misma forma en los grupos que se comparan?	¿Fue el seguimiento lo suficientemente prolongado y completo?
¿Ha sido el seguimiento suficientemente largo y completo?	¿Se utilizaron criterios objetivos y no sesgados para los resultados?
¿Es correcta la relación temporal?	¿Se hizo un ajuste por los factores pronósticos importantes?
¿Existe un gradiente en la relación dosis-respuesta?	

(Continúa)

Tabla 9. Valoración crítica de estudios cuantitativos CASP *(continuación)*

Ensayo clínico	Prueba diagnóstica
Aparte de la investigación experimental, ¿se ha tratado a los grupos de la misma forma?	
¿Cuáles son los resultados?	
¿Cuál es la magnitud del efecto?	¿Se presentan los cocientes de probabilidad *(likelihood ratios)* o los datos para calcularlos?
¿Cómo es la precisión de la estimación del efecto del tratamiento?	¿Cuál es la precisión de los resultados?
¿Son los resultados aplicables en mi medio?	
¿Se puede aplicar los resultados en mi medio o población local?	¿Serán satisfactorios en mi ámbito la reproducibilidad de la prueba y su interpretación?
¿Se tuvieron en cuenta todos los resultados de importancia clínica?	¿Es aceptable la prueba en mi paciente?
¿Los beneficios que obtener justifican los riesgos y los costes?	¿Modificarán los resultados de la prueba la decisión de cómo actuar?

Etiología	Pronóstico
¿Cuáles son los resultados?	
¿Cuál es la fuerza de la asociación entre la exposición y el resultado?	¿Cuál es la probabilidad del (los) evento(s) en un periodo de tiempo determinado?
¿Qué precisión tiene la estimación del riesgo?	¿Cuál es la precisión de las estimaciones?
¿Son los resultados aplicables en mi medio?	
¿Son aplicables los resultados a mi práctica clínica?	¿Son los pacientes del estudio similares a los mios?
¿Cuál es la magnitud del riesgo?	¿Conducen los resultados a seleccionar o a evitar un tratamiento?
¿Debo intentar detener la exposición?	¿Son útiles los resultados para tranquilizar o aconsejar a los pacientes?

teza que pueda haber tras cada decisión. Para estos casos, sería conveniente seguir las recomendaciones de Sackett[41] de dirigirnos a la última pregunta sobre la aplicabilidad en nuestros pacientes –sí/no– y tomar entonces una decisión.

Sirva como anécdota o chascarrillo algo que me ocurrió durante mis últimos cursos de carrera. Un profesor de Genética Molecular, cuyo nombre no desvelo por motivos obvios, amparándome en la nueva ley de protección de datos, siempre nos decía eso de "...en genética molecular, queridos alumnos, no existen blancos ni negros, sino una amplia gama de grises...". Esta frase me ha venido muy bien para tratar de explicar lo argumentado en el párrafo inmediatamente anterior. Eso sí, dejo para mí y para mi querido profesor una reflexión más profunda de por qué sus exámenes eran siempre de tipo test con cuatro o cinco opciones por pregunta, a cada cual más ambigua y con una sola opción correcta. ¿No habíamos quedado en que se trataba de una gama de grises?

El análisis crítico de un artículo científico basado en un estudio de investigación cualitativa (investigación de fenómenos no cuantificables, típicamente profunda y holística, me-

diante la colecta de material narrativo ilustrativo; p. ej., sexo, estado civil, estado de salud, gravedad de una enfermedad) tiene un enfoque muy similar al de los estudios cuantitativos. En los estudios cualitativos se busca comprobar la credibilidad y la fiabilidad de los resultados de la investigación **(tabla 10)**[5,6].

Asimismo, desde los centros impulsores de la lectura crítica se han ido desarrollando un gran número de guías diseñadas para cada tipo de estudio cuya finalidad es ayudar a los profesionales en su toma de decisiones, haciendo que se basen en la evidencia científica, reconociendo a la vez el grado de certeza y la fuerza de la inferencia en la que se fundamentan. Estas guías o listas de comprobación serán revisadas en profundidad en el capítulo 3 de este libro, y son las que figuran a continuación:

- STROBE para los estudios observacionales.
- CONSORT para los ensayos clínicos aleatorizados.
- QUORUM para los metaanálisis.

Tabla 10. Valoración crítica de estudios cualitativos CASP

¿Son válidos los resultados del estudio?	¿Cuáles son los resultados?	¿Son los resultados aplicables en mi medio?
1. ¿La investigación pretende estudiar un fenómeno en profundidad? 2. ¿Estaba bien definido el problema de investigación? 3. ¿La estrategia de muestreo es la adecuada para conseguir los objetivos? 4. ¿Fueron identificados con claridad y minuciosidad los métodos de recogida de datos, el escenario y los participantes? 5. ¿Se utilizó alguna técnica de triangulación? 6. ¿Se ha considerado de modo adecuado la interrelación entre investigadores y participantes?	1. ¿Son explícitos y fáciles de entender? 2. ¿Hay datos suficientes para justificar los resultados expuestos? 3. ¿Hay recomendaciones que pueden ser aplicables a la práctica clínica?	1. ¿Las características del grupo estudiado eran marginales? 2. ¿Contribuirán estos hallazgos a mejorar mi práctica asistencial? 3. ¿Puede servirme para conocer mejor el problema de un paciente concreto? 4. ¿Puede verse beneficiado el paciente por los hallazgos de este estudio?

Lectura crítica de un artículo científico
¿Cómo analizar un artículo publicado?
(1ª parte)

2

CRITERIOS GENERALES PARA LA LECTURA CRÍTICA DE ARTÍCULOS

2. CRITERIOS GENERALES PARA LA LECTURA CRÍTICA DE ARTÍCULOS

2.1. Evaluación preliminar

Una vez elaborada la pregunta clínica y encontrado el artículo (o artículos) en el que se trata de responder a la misma tras una búsqueda exhaustiva y sistematizada de la bibliografía, los siguientes pasos que plantea la medicina basada en la evidencia (MBE) son realizar una valoración crítica del documento encontrado y evaluar su aplicabilidad a nuestra práctica clínica[57]. Esta fase es fundamental ya que, a menudo, la calidad de los artículos científicos es deficiente, no se ajustan al problema clínico que se trata de resolver, tienen errores metodológicos que comprometen los resultados o estos son presentados de forma que limitan su correcta interpretación. Por ello, cuando caiga en nuestras manos un artículo que vayamos a utilizar en nuestra práctica clínica cotidiana, es aconsejable plantearse una serie de dudas razonables y ser sistemático en su lectura.

El método estructurado de pensamiento y trabajo que propone la MBE para resolver nuestras dudas derivadas de la práctica clínica habitual supone un cierto esfuerzo, pero sus resultados serán visibles a medio y largo plazo.

Para ello, es preciso conocer las técnicas y hábitos de aprendizaje desarrollados por el Evidence-Based Medicine Working Group de la Universidad McMaster (Canadá), y que ya han sido expuestos en el primer capítulo de este libro.

La valoración crítica de documentos científicos es una actividad dirigida a evaluar la validez de los trabajos y emitir así un juicio de valor. Esto va a depender del nivel de conocimiento de los métodos y técnicas en investigación clínica y del nivel de conocimiento que tenga el lector en el área de la investigación. Ante esta propuesta, mediante la lectura crítica, el lector tendrá una postura propia sobre lo expresado en el texto, descubriendo los supuestos implícitos, la idea directriz, los puntos fuertes y débiles de los argumentos, llegando a proponer incluso otros planteamientos que superen los del autor y reafirmando o modificando así su propia postura[58]. En este sentido, para Espinosa y Viniegra[59], la lectura crítica tiene tres componentes: *interpretar*, enjuiciar y proponer. Interpretar (en informes médicos) es la habilidad para reconstruir el contenido de un trabajo de investigación a partir de la identificación de cada uno de sus componentes: tipo y

diseño de estudio, tamaño y población de la muestra, variables e indicadores, procedimientos de medición, resultados (tablas, gráficos, etc.), y análisis final de los resultados. *Enjuiciar* permite diferenciar lo novedoso, relevante, pertinente, apropiado, válido y fuerte de lo que no lo es, en cada uno de los aspectos considerados en la interpretación. *Proponer* es formular alternativas para mejorar el artículo.

Cuando la crítica se aplica a los artículos de investigación clínica, la constante actividad cuestionadora del lector le va a permitir aclarar lo metodológicamente bien hecho y lo útil para resolver problemas de su entorno. Es relevante destacar la importancia que tiene el uso de la investigación clínica tanto para renovar conocimientos en la práctica médica como para resolver problemas de la misma. Además, debemos reconocer que la investigación puede y debe ser utilizada como herramienta de aprendizaje[60].

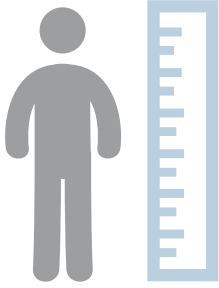
Por ello, la valoración crítica de un artículo de investigación en medicina es un proceso fundamental para el cual el clínico debe adquirir las destrezas y habilidades necesarias para juzgar si son válidos (próximos a la verdad y con rigor científico) y decidir sin son importantes (y, en consecuencia, valiosos en potencia para el lector en su condición de clínico) y aplicables (en la práctica clínica habitual).

Entendemos por *validez* la credibilidad que nos merece el artículo. Es decir, que cuando expresamos que un artículo es válido, estamos diciendo que creemos en él. La contribución fundamental de un experimento es la de ayudar a excluir las diferentes variables que podrían explicar un fenómeno. Pero un experimento determinado no excluye necesariamente todas las explicaciones alternativas posibles. Desde un punto de vista metodológico, el mejor experimento es aquel cuyo diseño excluye las explicaciones alternativas de los resultados. Idealmente, los resultados de un experimento tendrán solo una interpretación, que sería que la variable independiente (predictora) es la responsable de los cambios en la dependiente (resultado).

Un experimento resulta válido en la medida que los resultados pueden ser considerados el efecto verdadero, y en el grado que se generalizan fuera del experimento. Si los resultados de un experimento pueden ser atribuidos inequívocamente a la variable de predicción, se dice que el experimento es internamente válido. El grado en que permite la generalización se refiere a los aspectos relativos a la validez externa.

Surgen aquí dos conceptos que hemos de manejar a la hora de hacer la valoración preliminar de cualquier estudio, y son los de validez interna y validez externa.

- La **validez interna** se refiere al grado en que un experimento excluye las explicaciones alternativas de los resultados, es decir, al grado en que los resultados de un estudio son extraídos correctamente de los participantes del mismo. Como posteriormente veremos, los factores que afectan a la validez interna son los errores sistemáticos (sesgos) y los factores de confusión.



Si en un estudio se debe medir la estatura de diferentes personas en centímetros, utilizar una cinta métrica mal calibrada implica que los datos son de por sí falsos.

Figura 6a. Ejemplo de validez interna.

Si en el estudio anterior se utilizan jugadores de baloncesto para calcular la estatura media (utilizando una cinta métrica correctamente calibrada), es más que probable que estos resultados, ciertos en la muestra, no sean extrapolables al resto de la población.

Figura 6b. Ejemplo de validez externa.

- La **validez externa** se refiere al grado en que los resultados de un estudio son generalizables a otras poblaciones distintas a la estudiada. Para poder generalizar los resultados de un estudio, este ha de ser válido internamente, y ha de haber consistencia (precisión) en los resultados obtenidos a partir de estudios diferentes llevados a cabo, utilizando poblaciones diferentes.

Podemos concluir entonces que, mientras que la validez interna dirige la pregunta inicial referente a si un experimento concreto ha demostrado una relación inequívoca, la validez externa hace una pregunta más amplia, referente al grado

con que pueden generalizarse los resultados de un experimento **(figuras 6a** y **b)**. Además, ambos conceptos no pueden trabajar opuestamente. Si la validez interna es máxima, significa que el trabajo dispone de controles muy exhaustivos que no serían aplicables a otros grupos, sujetos o lugares donde tales controles no existen. No obstante, es preferible asegurarse de que un estudio está libre de sesgos (internamente válido) y se lleve a cabo en condiciones muy controladas, sacrificando la validez externa, y más tarde desarrollar experimentos con menor control, estableciendo una validez externa superior.

Otra de las prioridades de la MBE es llegar a valorar la importancia o relevancia clínica de un artículo, incluso mucho más allá de la cuestión estadística. Este paso nos remite al conocimiento de los temas metodológicos de la evidencia científica, lo que implica al menos un pequeño esfuerzo por parte del clínico, necesario para poder entender y juzgar lo que se publica. En el capítulo 4 recurriremos a conceptos metodológicos y epidemiológicos fundamentales (*odds ratio*, riesgo relativo, índice Kappa, intervalo de confianza) para la etapa de valoración crítica en MBE y una comprensión correcta de los resultados.

Tras demostrar que la mejor evidencia externa encontrada en la literatura científica es válida (verdadera y precisa) e importante, nos encontramos ante la pregunta decisiva y fin último de la MBE: ¿se puede integrar esta evidencia científica con

nuestra maestría clínica e incorporarla a la práctica clínica con nuestros pacientes? Este paso puede tratarse del interés que tengamos en un paciente concreto, o extrapolarlo a la generalidad de los pacientes con un fin concreto (intervención diagnóstica o terapéutica), lo que podría constituir las bases para el desarrollo de guías de práctica clínica.

Hasta aquí, podemos decir que la valoración crítica de la literatura médica es un paso difícil, pero no imposible de la MBE. Aunque los criterios utilizados están suficientemente acreditados para ayudar a los profesionales de la medicina en el mejor uso de la bibliografía, permiten un cierto margen de interpretación subjetiva, y si se aplican muy estrictamente pueden hacer que esta práctica termine en un simple recetario.

Está claro que no todos los alergólogos, pediatras, cirujanos, etc. tienen que ser expertos en la metodología de la medicina basada en la evidencia, pero sí deben ser consumidores inteligentes de la bibliografía y de argumentos científicos. Para ello será necesario empezar por hacer una descripción más detallada de lo que se consideran los criterios generales de la lectura crítica.

2.2. Las preguntas más frecuentes

La valoración crítica de un artículo científico va a ser diferente en función del tipo de pregunta clínica y del diseño elegido en el estudio evaluado. Existen criterios de valoración comunes a cualquier estudio que habrían de responder a las siguientes cuestiones básicas esquematizadas en la **figura 7**.

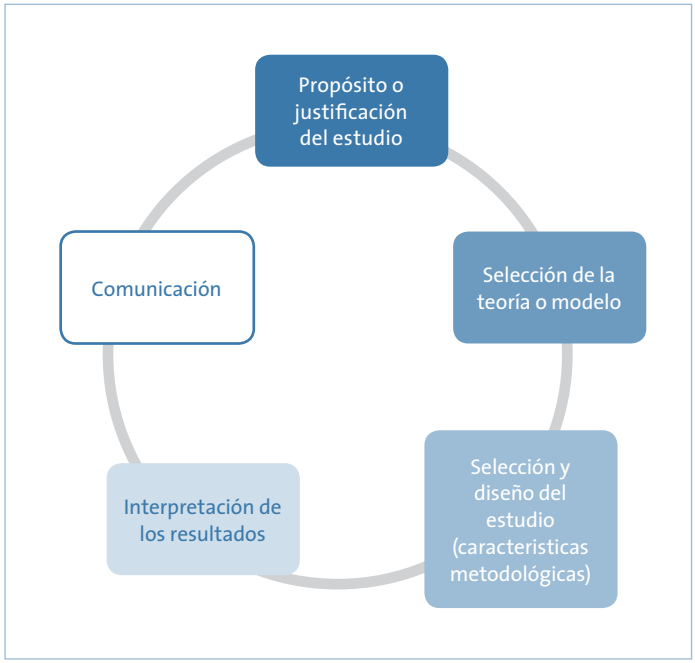


Figura 7. Criterios generales que seguir durante la evaluación crítica de un artículo.

Conforme avancemos en la lectura de este capítulo, el seguimiento de este esquema nos lleva a valorar, entre otras cosas, el grado de adecuación del diseño y de la población estudiada a la pregunta de investigación, el tamaño de la muestra –suficiente o insuficiente–, la homogeneidad de los grupos comparados al inicio del estudio, el seguimiento uniforme y completo de los sujetos de estudio y la medición y análisis apropiados de los resultados e interpretación adecuada de los mismos.

La evaluación de la validez es, en la práctica, una evaluación del diseño del artículo y de su desarrollo. La estrategia de evaluación de la validez consiste en responder una serie de preguntas relacionadas con el diseño y el desarrollo de la investigación, y de cuya respuesta (positiva o negativa) dependerá que aceptemos o no su validez como evidencia científica.

Así, las preguntas que deberemos plantearnos son las siguientes:

- ¿Cuál es la pregunta de investigación?
- ¿Se trata de un estudio original?
- ¿Por qué se hizo el estudio?

- ¿Qué tipo de estudio llevaron a cabo los investigadores?
- ¿Es el diseño del estudio adecuado para la investigación?
- ¿Se han incluido estrategias para optimizar la validez y precisión del estudio?
- ¿Es adecuada la población estudiada?
- ¿Son adecuadas las características metodológicas del estudio?

2.2.1. ¿Cuál es la pregunta de investigación?

Un artículo científico original trata de contar la historia de una investigación. Su valoración crítica tendrá que comenzar necesariamente juzgando si el estudio trata de contestar a nuestra pregunta de investigación. Esta cuestión debería haberse resuelto en las fases de búsqueda y selección del artículo. Sin embargo, debe ser comprobada antes de dedicar nuestro tiempo a un análisis más detallado. Para ello, nos fijaremos en el tipo de población estudiada y de intervención terapéutica o diagnóstica evaluada, el diseño del estudio y los criterios empleados para la medición de los resultados[61,62].

La pregunta de investigación es el punto clave de cualquier estudio, ya que nos dice el valor que tiene esa investigación en el marco del conocimiento actual y condiciona el diseño del

estudio[62]. En el primer capítulo de este libro (véase el apartado 1.2.1) ya enumeramos las características que debe reunir una buena pregunta clínica. Hulley et al.[1] utilizan las siglas FINER, que recogen las cinco características básicas de una pregunta de investigación: factible *(feasible)*, interesante *(interesting)*, novedosa *(novel)*, ética *(ethical)* y relevante *(relevant)*.

El hecho de que la pregunta del estudio sea *factible* se refiere a la capacidad de los investigadores de contestar a la pregunta que se plantearon. Esto es, si era posible incluir el número de pacientes necesarios para contestar a la pregunta del estudio, si el conocimiento y experiencia de los investigadores era el adecuado; en definitiva, si los recursos disponibles eran suficientes y adecuados como para llevar a cabo el estudio.

La pregunta del estudio debe ser *interesante* para los investigadores que van a participar en él, y también debe ser «relevante» para la comunidad científica, bien sea porque su respuesta suponga un avance en el conocimiento científico, bien porque sirva de guía para la realización de futuras investigaciones o sea de ayuda para la toma de decisiones clínicas o de salud pública.

2.2.2. ¿Se trata de un estudio original?

Otro aspecto clave de la pregunta de investigación es la *novedad*. En investigación básica, a menos que haya indicios de que puedan encontrarse nuevos resultados o controversias

sobre lo que se creía demostrado, no tendría ningún sentido investigar de nuevo un mismo fenómeno.

Pero esta reflexión tiene sus matices cuando se habla de investigación clínica. En investigación clínica, en la que participan personas, el hecho de evitar la investigación redundante está estrechamente ligado a la *ética* del estudio, ética que sirve de contrapunto a los intereses del investigador y de la comunidad científica. Según la Declaración de Helsinki[63], en la investigación con seres humanos "el bienestar de la persona que participa en la investigación debe tener siempre primacía sobre todos los otros intereses". Por ello, la investigación clínica "está sujeta a normas éticas que sirven para promover el respeto a todos los seres humanos y para proteger su salud y sus derechos individuales", entre los que se incluyen "el derecho a la autodeterminación, la intimidad y la confidencialidad de la información personal".

Ahora bien, solo una pequeña parte de la investigación médica es completamente novedosa, y análogamente, solo una pequeña parte repite exactamente un trabajo previo. La gran mayoría de la investigación nos dirá, en el mejor de los casos, que una hipótesis determinada es, probablemente, un poco más correcta o un poco más incorrecta de lo que era antes de agregar esta nueva información. Por ello, es perfectamente válido llevar a cabo un estudio que no sea completamente original. De hecho, la ciencia del metaanálisis se basa en la aparición de más literatura destinada a responder la misma pregunta de un modo similar.

De este modo, la pregunta práctica que hay que realizar no es ¿alguien ha hecho antes un estudio similar? Lo correcto es preguntarse ¿agrega esta investigación un conocimiento nuevo al problema en estudio?

Es decir, se trata de averiguar si se trata de un estudio más grande, o más largo, o más sustancial que los anteriores; con una metodología más rigurosa, o si contiene particularmente alguna crítica específica a los métodos empleados en la metodología de los estudios previos.

2.2.3. ¿Por qué se hizo el estudio?

Novedad y relevancia son puntos claves de la investigación clínica. Durante la lectura crítica del artículo, nos ha de quedar claro qué hueco del conocimiento se pretende rellenar contestando a la pregunta clínica –*novedad*– y por qué era importante rellenar ese hueco –relevancia–. Se trata, en definitiva, de establecer la *razón de ser del estudio*.

La razón de ser del estudio debe preceder a la pregunta de investigación, es el camino que llevará al lector hacia ella. En otras palabras, con la razón de ser del estudio debemos aportar al lector la información necesaria para entender la pregunta o hipótesis del mismo. En ese camino iremos de lo general (an-

tecedentes del problema) a lo particular (la pregunta o hipótesis), es decir, de lo conocido a lo desconocido **(figura 8)**. Ante la investigación de una enfermedad o un problema clínico, el lector podrá hacerse una primera idea de la importancia del problema tratado a partir de los datos epidemiológicos o de impacto (calidad de vida, mortalidad, morbilidad, etc.). Inmediatamente después, se ha de discutir lo conocido hasta ahora sobre el tema de investigación, detallándose qué hallazgos previos son la base de la hipótesis que se plantea, de forma que se construya el marco en el que encaja la pregunta del estudio. Por ejemplo, en un ensayo clínico sobre un nuevo medicamento se debe dejar claro qué tratamientos hay actualmente disponibles y sus limitaciones en cuanto a eficacia, tolerabilidad e impacto en la calidad de vida del paciente. Si el medicamento pertenece a una clase farmacológica nueva para el tratamiento de esa enfermedad, deberemos proporcionar, además, las bases fisiopatológicas que llevaron a pensar en el uso de dicho medicamento.

De lo conocido se ha de pasar a lo desconocido, de modo que en el artículo se ha de indicar qué es lo que no se conoce aún sobre el asunto por investigar (p. ej., la eficacia de un nuevo medicamento medida en un ensayo clínico), haciendo evidente, implícita o explícitamente, qué es lo novedoso del estudio (como primer estudio sobre la eficacia de un nuevo medicamento o estudio en una población distinta o más representativa).

Figura 8. Esquema de presentación del porqué de un estudio en un artículo científico (modificado de Villademoros y Hernando[62]).

El colofón de la razón de ser de un estudio es la pregunta de investigación o hipótesis del estudio, entendida como el enunciado del interrogante específico que el investigador desea responder para resolver un problema de investigación.

2.2.4. ¿Qué tipo de estudio llevaron a cabo los investigadores?

Una de las etapas más importantes de la lectura crítica de un artículo de investigación clínica consiste en saber reconocer el tipo de estudio que se ha desarrollado previamente a la publicación.

Por diseño de un estudio entendemos el conjunto de procedimientos, métodos y técnicas mediante los cuales el investigador selecciona a los pacientes, reúne una información, la analiza e interpreta los resultados. Tras la fase de establecimiento del concepto del problema de investigación, la siguiente que ha de cubrir el investigador es la de decidir cuál es el diseño que le va a permitir contestar a nuestra pregunta de investigación de la forma más eficaz posible.

Por tanto, antes de pasar a evaluar el diseño del estudio en cuestión, nos va a resultar útil encuadrar nuestra pregunta de investigación en uno (o varios) de estos enfoques o cate-

gorías. De este modo la revisión del trabajo será más fácil si disponemos de una lista más reducida de los diseños posibles de investigación, utilizándola con un criterio común, cualquiera que sea el artículo que valoremos: la elección del tipo de estudio por parte de los autores siempre ha de ser aquella que presente menos problemas de validez y sea al mismo tiempo, compatible con la disponibilidad de recursos. Como veremos a continuación, estos enfoques no son excluyentes, ya que por un lado contienen diseños comunes y, por otro, pueden concurrir en un mismo estudio.

Los estudios clínicos de investigación se pueden clasificar, atendiendo a diversos criterios. Clásicamente, según su finalidad –*analíticos y descriptivos*–, el control de la asignación de los factores del estudio –*experimentales y observacionales*– y la cronología de los hechos –*prospectivos y retrospectivos*–, y la secuencia temporal –*longitudinales y transversales*–. La clasificación de estos diseños nos permitirá tener una idea rápida de las características del método que se utiliza **(tabla 11)**.

Tipos de estudios según su finalidad

Los **estudios descriptivos** son aquellos que describen la frecuencia y las características más importantes de un problema de salud[64]. El objetivo de los estudios descriptivos es describir una enfermedad o característica en una o varias poblaciones determinadas. Por tanto, son útiles cuando se conoce poco acerca de lo que queremos estudiar y normalmente sirve de inicio a futuras investigaciones.

Tabla 11. Clasificación de los tipos de estudios en investigación clínica

Tipos de estudios por su finalidad	
Analíticos	***Descriptivos***
• Observacionales – Estudios de cohortes – Estudios de casos y controles • Intervención – Ensayo clínico – Ensayo de campo – Ensayo comunitario	• En poblaciones – Estudios ecológicos • En individuos – Estudios de incidencia – Series de casos – Prevalencia (transversales)
Tipos de estudios por el control de asignación de factores	
Experimentales	***Observacionales (no experimentales)***
• Ensayo clínico • Ensayo de campo • Ensayo comunitario de intervención	• Estudios ecológicos • Estudios de prevalencia • Estudios de casos y controles • Estudios de cohortes o de seguimiento
Tipos de estudios por la cronología de los hechos	
Prospectivos	***Retrospectivos***
Tipos de estudios por la secuencia temporal	
Longitudinales	***Transversales***

Podemos distinguir diferentes tipos de estudios descriptivos según el diseño de los mismos:

- **Estudios ecológicos:** estos estudios no utilizan la información del individuo de una forma aislada, sino que usan datos agregados de toda la población. Describen la enfermedad en la población con relación a variables de interés (la edad, la utilización de servicios, el consumo de alimentos, de bebidas alcohólicas, de tabaco, la renta per cápita, etc.). En muchas ocasiones, estos estudios son el primer paso en la investigación de una posible relación entre una enfermedad y una exposición determinada. (p. ej., correlacionar la incidencia de asma alérgico por exposición a *Aspergillus* sp.).
- **Estudios de incidencia:** son estudios de casos individuales, en los que la unidad de análisis es el individuo. En ellos se describe una serie de variables en una población determinada que es seguida a lo largo del tiempo.
- **Series de casos:** estos estudios describen la experiencia de un paciente o un grupo de pacientes con un diagnóstico similar en los que todos ellos tienen la misma enfermedad o reciben el mismo tratamiento. Con frecuencia se describe una característica de una enfermedad o de un paciente que sirve para generar nuevas hipótesis. Muchas veces documentan la presencia de nuevas enfermedades o efectos adversos y, en este sentido, sirven para mantener una vigilancia epidemiológica.

Se considera **analítico** todo estudio que evalúa una presunta relación causa-efecto. El presunto agente puede ser tanto un factor etiológico como un tratamiento o intervención para prevenir o mejorar una situación clínica. Son diferentes los tipos de estudios analíticos: por un lado, están los **estudios de intervención**, en los que el investigador tiene la facultad de manipular las condiciones de la investigación. Y, por otro, los **estudios observacionales**, cuya característica fundamental es que el investigador no posee la facultad de decidir qué sujetos van a estar expuestos a la influencia de un cierto factor bajo estudio, aun cuando sea posible decidir cuándo y dónde se ha de realizar la investigación[65,66]. Es decir, aquí el investigador se limita a observar, medir y analizar determinadas variables, sin ejercer un control directo de la intervención.

La selección de los participantes en el estudio se puede llevar a cabo de acuerdo con la exposición, con el evento o sin considerar ninguna de estas características de los sujetos elegibles para el estudio[67]. Son los propios sujetos de estudio o el profesional sanitario dentro de su práctica asistencial habitual los que pueden elegir la exposición, o incluso puede venir impuesta si interesara incluir personas de una raza o un género determinado.

A partir de aquí utilizaremos el término "exposición" para referirnos a la variable en estudio. Exposición como un término de significado amplio, que puede abarcar desde la exposición a una bacteria o a una sustancia tóxica hasta la exposición a un suplemento nutricional, una vacuna, un programa de salud o un estilo de vida. Utilizaremos el término evento para referirnos a la variable respuesta o cambio que se espera detectar con relación a la exposición. Se utiliza también de manera amplia y puede referirse tanto a la ocurrencia de una enfermedad, como al cambio de estado o cambio promedio en una variable continua; por ejemplo, cambio en las concentraciones de colesterol sérico o seroconversión después de ser aplicada una vacuna.

Si se asume que la enfermedad no ocurre al azar es obvio, entonces, que cuando en un individuo o grupo humano se presenta una enfermedad esa ocurrencia ha sido determinada por el contacto previo con uno o varios factores capaces de dar origen a ese estado. Al contrario, puede suponerse que una población libre de esa enfermedad no estuvo expuesta a los factores condicionantes o que, de algún modo, estuvo protegida ante la exposición.

Pero esto tampoco resulta tan fácil, puesto que en la mayoría de los casos la enfermedad se distribuye en la población con relación a un grupo de variables que participan del origen de tal enfermedad, pero también frente a un gran número de

variables que van a hacer más difícil el análisis de la relación causa-efecto que estamos buscando.

Esta es la razón fundamental de lo sofisticado de algunos de estos diseños observacionales:

- **Estudios de cohortes:** en este tipo de estudio los individuos son identificados en función de la presencia o ausencia de exposición a un determinado factor, para determinar después la ocurrencia del evento. En este momento todos están libres de la enfermedad de interés y son seguidos durante un periodo de tiempo para observar la frecuencia de aparición del fenómeno que nos interesa. Si al finalizar el periodo de observación la incidencia de la enfermedad es mayor en el grupo de expuestos, podremos concluir que existe una asociación estadística entre la exposición a la variable y la incidencia de la enfermedad. Es el tipo de estudio epidemiológico de observación que más se acerca al experimento. Su característica fundamental es que se inicia con la detección de una supuesta causa y luego sigue a una población a través del tiempo hasta determinar si se presenta el efecto.
- **Estudios de casos y controles:** en contraste, cuando se selecciona a los participantes a partir del evento de estudio, es decir, se elige de manera independiente un grupo de sujetos que tienen el evento de interés (casos) y un grupo de sujetos que no lo tienen (controles) y en estos grupos se determina la exposición, entonces hablamos de un es-

tudio de casos y controles. Sobre la base de observaciones clínicas, o a través de análisis especiales, los estudios de casos y controles revisan situaciones de exposición a factores sospechosos comparando grupos de individuos enfermos (casos) con grupos de individuos sanos (controles), según la frecuencia observada en cada grupo de la exposición previa al factor sospechoso. La relación entre uno o varios factores relacionados con la enfermedad se examina comparando la frecuencia de exposición a este u otros factores entre los casos y los controles[68,69]. Si a través del análisis estadístico de los datos puede establecerse que la presencia del factor en estudio es significativamente más frecuente entre los casos que entre los controles puede hablarse, entonces, de una asociación entre la enfermedad y dicho factor.

No es objetivo de este libro detallar los pormenores metodológicos de estos estudios, pero sí consideramos importante aclarar que el rigor en la selección de casos y controles para asegurar la comparabilidad es el punto fundamental para el éxito del estudio. Los grandes temas que se deben abordar al realizar un estudio de casos y controles son después de la definición de caso, la selección de los controles y las fuentes de información sobre la exposición y la enfermedad. Resumiremos así las condiciones que han de cumplirse en estos estudios respecto a los casos y controles (tabla 12).

Tabla 12. Condiciones que se han de cumplir en los estudios de casos y controles

Respecto a los casos	Respecto a los controles
1. Establecer de forma clara y explícita la definición de la enfermedad y los criterios de inclusión 2. Los casos deben ser incidentes ya que los casos prevalentes: – Cambian sus hábitos en relación con la exposición – Pueden ser los supervivientes de casos incidentes y la supervivencia puede estar relacionada con la exposición	1. La función del grupo control es estimar la proporción de exposición esperada en un grupo que no tiene la enfermedad 2. Los controles deben ser representativos de la población de donde provienen los casos. Los casos y los controles no deben entenderse como dos grupos de dos poblaciones distintas, sino como dos grupos que proceden de una misma población 3. Los controles deben ser comparables a los casos y tener la misma probabilidad de haber estado expuestos

Tipos de estudios según la dirección temporal

De acuerdo con el número de mediciones que se realiza en cada sujeto de estudio para medir la ocurrencia del evento o los cambios en la variable de exposición a lo largo del tiempo, los estudios se definen como **longitudinales** (cuando se realizan al menos dos mediciones: la medición basal para determinar el estado inicial y una subsecuente para determinar la ocurrencia del evento) y como **transversales** (cuando se realiza una sola determinación en los sujetos de estudio y se evalúan de manera concurrente la exposición y el evento de interés).

- **Estudios longitudinales:** la discusión sobre el significado del término longitudinal la resumió Chin[70] en 1989: para los epidemiólogos es sinónimo de estudio de cohortes o seguimiento, mientras que para algunos estadísticos implica mediciones repetidas. Según la definición dada se considera longitudinal si las observaciones se refieren a dos momentos en el tiempo, aun cuando la información se recoja de forma simultánea, como ocurre en los estudios de casos y controles. La palabra longitudinal denota la existencia de un intervalo de tiempo entre la exposición y el comienzo de la enfermedad[71]. Bajo esta acepción, el estudio de casos y controles, que es una estrategia de muestreo para representar esa experiencia de la población a lo largo del tiempo, también sería un estudio longitudinal.

Conviene matizar que la visión estadística de un estudio longitudinal parte de un análisis de datos diferente, aprovechándose de las mediciones repetidas en los sujetos a lo largo del tiempo, permitiendo una inferencia, además de poblacional, a nivel individual en los cambios de un proceso a lo largo del tiempo o en las transiciones entre diferentes estados de salud y la enfermedad[72]. Los estudios de cohortes son los que resumen la exposición y la enfermedad en los grupos en los que se comparan, por ejemplo, la frecuencia del asma alérgica en fumadores y no fumadores. La inferencia proporcionada por estos estudios se refiere a medias poblacionales. Un ejemplo de este tipo de estudio puede ser el *Nurses' Health Study*, con más 120.000 enfermeras en 11 Estados de EE.UU., en el que la exposición acumulada a los anticonceptivos orales se valoró como factor de riesgo del cáncer de mama[73].

- **Estudios transversales:** estudian simultáneamente la exposición y la enfermedad en una población bien definida en un momento determinado. Esta medición simultánea no permite conocer la secuencia temporal de los acontecimientos y no es, por tanto, posible determinar si la exposición precedió a la enfermedad o viceversa. Valga el siguiente ejemplo para explicarlo: si realizamos un estudio en el que determinamos la asociación en un momento dado entre la obesidad infantil y el dolor de rodilla, podremos saber si ambas patologías están relacionadas, pero no podremos determinar si los niños con dolor de

rodillas son más obesos (porque andan menos), o si los niños con obesidad padecen más dolor de rodillas (porque sobrecargan más la rodilla), ya que desconocemos la secuencia temporal.

También se le denomina estudio de encuesta y se distingue porque se indaga sobre la presencia de la exposición y la ocurrencia del evento una vez conformada la población en estudio, y porque solo se hace una medición en el tiempo en cada sujeto de estudio[74].

Un estudio de corte transversal es sinónimo de un estudio de prevalencia, es decir, un estudio en el que se describe la magnitud de un problema de salud en la población (una enfermedad, un factor de riesgo).

Una publicación ejemplo de este tipo de estudios es: 'Evolución temporal de la prevalencia de síntomas de rinitis alérgica en escolares de 13-14 áreas españolas en 1993-1994 y 2001-2002 según el Estudio Internacional sobre Asma y Alergias en la Infancia (ISAAC)[75]'.

Para la consecución de este estudio, Arnedo-Pena et al. aplicaron un cuestionario sobre síntomas de rinitis cumplimentado a sujetos en edad escolar (n = 30.046)

procedentes de 9 ciudades españolas. Las prevalencias de rinoconjuntivitis y de alergia nasal experimentaron distintos patrones de evolución por factores como sexo, tiempo estacional y centro. De ahí que el estudio de los factores de riesgo en el ámbito local sea de importancia para su prevención.

Esta información es de gran utilidad para valorar el estado de salud de una comunidad y determinar sus necesidades. Asimismo, sirven como todos los estudios descriptivos, para formular hipótesis etiológicas.

En términos de causalidad existe una diferencia importante entre estos dos tipos de estudios. En los longitudinales es posible verificar que la exposición antecede a la ocurrencia del episodio, con lo que se cumple el principio temporal de causalidad –la causa antecede al efecto–. Por el contrario, en los transversales resulta imposible verificar este tipo de relaciones cuando se estudian exposiciones que cambian con el tiempo. Los estudios transversales sí pueden proporcionar información valiosa cuando se estudian factores que no varían (como el sexo y la carga genética) o exposiciones únicas que no cambian con el tiempo (por ejemplo, el caso de la población expuesta a una bomba atómica).

Tipos de estudios según la cronología de las observaciones

De acuerdo con el tiempo de inicio del estudio con relación a la cronología de los hechos, los estudios se dividen en **prospectivos** y **retrospectivos**. El criterio de temporalidad en la ocurrencia del episodio se utiliza para distinguir entre los estudios retrospectivos y prospectivos. El punto de referencia para esta clasificación es la ocurrencia del evento de interés (la variable respuesta). Si al inicio del estudio, el evento investigado ya ocurrió y el investigador planea reconstruir su ocurrencia en el pasado utilizando registros o entrevistando a los mismos sujetos de estudio, se considera que el estudio es retrospectivo. Si la ocurrencia del evento se registra durante el estudio, es decir, si los sujetos de estudio están libres del evento de interés al iniciar su participación en el estudio, el diseño se considera de tipo prospectivo.

- **Estudios retrospectivos:** su objetivo principal es probar alguna hipótesis planteada sobre la etiología de una enfermedad, es decir, que estos se dedican al análisis de una presunta relación entre algún factor y el desarrollo de cierta enfermedad. Los estudios retrospectivos se hacen a partir de observaciones clínicas, o de análisis especiales, en los que se revisan situaciones de exposición a factores sospechosos, comparando grupos de individuos enfermos (casos), con grupos de individuos sanos (controles). A partir de la frecuencia observada en cada uno de los

grupos expuestos al factor en estudio se realiza un análisis estadístico. Si tras el análisis estadístico de los datos, se logra determinar la presencia del factor en estudio, entonces se puede establecer una asociación entre dicho factor y la enfermedad. Básicamente, se puede decir que este tipo de estudios busca las causas a partir de un efecto que ya se presentó. Los estudios retrospectivos parten de un efecto y regresan a buscar la causa.

- **Estudios prospectivos:** tienen como característica más importante la de iniciarse con la exposición de una supuesta causa, y luego seguir a través del tiempo a una población definida hasta determinar o no la aparición del efecto[64]. Este tipo de estudio es muy utilizado en epidemiología. Cuando se realiza un estudio prospectivo, una vez planteada la hipótesis, se define la cohorte que participará en la observación (grupo de la misma edad, individuos que practican una profesión e incluso a todo un sector o una comunidad). Se dividen los grupos en función de su exposición o no a un supuesto factor causal, el cual puede ser un factor individual, ambiental o ambos. La población sujeta al estudio (expuesto y no expuesto) se observa a través del tiempo. El seguimiento se realiza mediante la aplicación de cuestionarios, exámenes clínicos periódicos, y seguimiento de registros especiales y rutinarios, entre otros. La duración del seguimiento varía –meses, años– hasta que se presente el efecto postulado en la hipótesis.

Aunque es difícil definir las indicaciones de un estudio prospectivo, su mayor utilidad radica en la prueba de hipótesis sobre la etiología de enfermedades más o menos comunes como la hipertensión arterial o la enfermedad coronaria en adultos, o la dermatitis atópica en el paciente pediátrico.

En general, podríamos afirmar que los estudios prospectivos tienen mayor puntaje en la escala de causalidad, dado que en este tipo de estudios se pueden diseñar instrumentos para la medición y registro del episodio que aseguren la calidad de las mediciones. En cambio, en los estudios retrospectivos la calidad de medición y registro del evento depende con frecuencia de instrumentos que no han sido diseñados de manera expresa para observar el evento en cuestión ni para responder a los objetivos de la investigación.

Los estudios que incluyen eventos que ocurrieron antes de iniciar la investigación y eventos evaluados de manera prospectiva son referidos en algunos textos como **mixtos** o **ambispectivos**.

Tipos de estudios según el control de asignación a los factores de estudio: los estudios experimentales

En el **estudio experimental**, el investigador tiene el manejo de la variable independiente, y puede manipular las condiciones de la investigación de manera intencionada. Este tipo de diseño se utiliza cuando se tiene que probar la eficacia y la seguridad de un nuevo fármaco, cuando se desea conocer la protección que brinda una vacuna, cuando se investiga la etiología de una enfermedad o para la evaluación de actividades de planificación y programación sanitarias.

Aquellos casos en que la epidemiología experimental introduce un elemento nuevo, o lo suprime, en el devenir natural de un proceso que afecta a una colectividad entera y sigue los resultados que se derivan como consecuencia de tal intervención, reciben el nombre de experiencias en condiciones no controladas (ausencia de grupo control). Un ejemplo de ellas serían las vacunaciones masivas en periodos de epidemia.

Es decir, el gran control que se tiene sobre el diseño facilita la interpretación de las asociaciones como causales.

La esencia de los estudios experimentales consiste en que el propio investigador decide qué individuos serán sometidos al tratamiento en prueba (grupo experimental) y quiénes estarán en el grupo comparativo (grupo control o de contraste). Es justamente esta propiedad la que distingue los ensayos clí-

nicos experimentales de los estudios observacionales (de cohortes), ya que en estos el propio paciente o su médico decide quién se somete (y quién no) a la maniobra en evaluación.

Para el alergólogo y cualquier otro especialista, los estudios experimentales son de gran interés. En palabras del propio Sackett[76]: "... los ensayos experimentales le van a permitir al clínico poder realizar inferencias causales en medio de tanta incertidumbre que rodea la práctica clínica ya sea en actividades de prevención, de diagnóstico o terapéuticas".

Esta decisión obedece a múltiples razones, algunas estrechamente ligadas al pronóstico de la enfermedad. Puede ser dejada al azar y, de manera aleatoria, los sujetos participantes en el estudio se dividen en dos grupos: el experimental, formado por los receptores del nuevo factor introducido por el investigador, y el control, constituido por los restantes participantes[66]. La aleatorización en la distribución de la intervención es lo que define un **estudio experimental como puro o verdadero**, comúnmente denominado **ensayo clínico aleatorizado**. Su característica distintiva es que los sujetos participantes son distribuidos de forma aleatoria entre los grupos de intervención y el que actúa como control.

Si, por el contrario, no hay aleatorización en la distribución del factor que se analiza y es el investigador –influido o no por las características de la experiencia– el que decide quiénes recibirán la intervención, el estudio es definido como **cuasiexperimental**. Estos son los **ensayos comunitarios**, más frecuentes en la epidemiología ambiental. En este tipo de estudios, el investigador u otras circunstancias no aleatorias determinan qué sujetos recibirán la medida que va a evaluarse y quiénes no.

- **Ensayo clínico:** Sir Arthur Bradford Hill[77], uno de los padres del ensayo clínico como lo conocemos hoy día, lo definió como "un experimento cuidadosa y éticamente diseñado con el fin de poder contestar a preguntas concretas formuladas previamente". En España, las publicaciones que regulan el ensayo clínico matizan un poco más y lo definen como una experiencia para la evaluación de una sustancia o medicamento, a través de su administración a los seres humanos con los siguientes fines: *a)* poner de manifiesto sus efectos farmacodinámicos; *b)* establecer su eficacia para una indicación terapéutica, profiláctica o diagnóstica determinada, y *c)* conocer el perfil de sus reacciones adversas y establecer su seguridad.

Un ensayo clínico es una evaluación experimental de un producto, sustancia, medicamento, técnica diagnóstica o terapéutica que a través de su aplicación a seres humanos pretende valorar su eficacia y seguridad. Recientemente, y por el agravamiento de la situación económica internacional, se ha propuesto que en la práctica médica diaria se empleen solo los medicamentos y procedimientos que hayan demostrado mayor eficacia y eficiencia sobre la salud de la población con el menor coste posible.

En la práctica médica actual un ensayo clínico aleatorio controlado de una terapéutica contra otra es la norma aceptada por la cual se juzga la utilidad de un tratamiento.

Estas definiciones comienzan a establecer que el ensayo clínico tenga que ser controlado, aleatorizado o enmascarado. La característica principal del ensayo clínico radica en la aleatorización: la asignación al azar de los sujetos para la formación de un grupo experimental y de un grupo control. Es el método electivo ideal para alcanzar la comparabilidad entre los grupos, ya que el investigador no selecciona los sujetos que van a pertenecer a cada grupo. Para ello ha de ser enmascarada (desconocida) para los investigadores que aplicarán la intervención. La responsabilidad de la misma depende del comité

de monitorización y ha de ser realizada por una persona ajena a la recogida de información y a la práctica de la intervención.

En definitiva, la aleatorización intenta conseguir que todos los factores de confusión se distribuyan por igual en los grupos de estudio, contribuyendo a que la validez interna del mismo sea lo más alta posible. Ante el problema de inferir los resultados a una población de referencia, los principales recursos metodológicos incluyen el apareamiento y la estratificación del universo de estudio en función de aquellas variables cuya influencia sobre el resultado previsto es conocida, por ejemplo: edad, género, raza, gravedad del padecimiento, etc. Así, al formar dos grupos al azar, se tiene la garantía de que ambos grupos son comparables para todos los factores con excepción del que se desea estudiar[7,8].

Por tratarse de una investigación clínica conducida en seres humanos, será necesario tomar en consideración una variedad de argumentos éticos. A pesar de partir de una serie de evidencias, siempre hay lugar para la incertidumbre, pues los individuos pueden sufrir un daño en lugar de un beneficio. Además, puede ocurrir que los recursos disponibles sean limitados, de modo que en caso de que se vaya a producir un beneficio, la pregunta es ¿quién lo va a recibir? Metodológicamente, la respuesta más clara es la de asignación al azar, ya que las presiones por recibir un procedimiento aún en experimentación pueden ser muy intensas y afectar al desarrollo del estudio.

Por lo general, los ensayos clínicos se proponen cuando se ha identificado alguna alternativa para la prevención o el tratamiento de una enfermedad, o bien, cuando se tienen evidencias de que una simplificación de la atención médica que normalmente se ofrece no acarrearía perjuicios al individuo y sí, en cambio, permitiría un ahorro de los recursos del sistema de salud[64].

¿Por qué y cuándo son necesarios los ensayos clínicos? ***Cuando nos enfrentamos a la posibilidad de que un nuevo tratamiento cambie la historia natural de una enfermedad y, por ende, la vida de muchos pacientes, hemos de plantearnos la necesidad de disponer de un conjunto completo de observaciones sistematizadas que permitan conocer el efecto en cuestión, antes de prescribirlo por sistema a los enfermos. En ocasiones, conocer la patogenia de una determinada enfermedad, así como los mecanismos de acción de un determinado fármaco nos puede llevar a probarlo, aunque desde luego, sin garantía de éxito. El uso de citarabina –inhibidor in vitro de los virus ADN como el virus del herpes– no rindió resultado alguno al ser usada directamente en pacientes.***

Otras dos fuentes de ideas sobre el posible valor de una cierta terapia suelen ser: a) ***las observaciones empíricas de clínicos perspicaces; así, por fortuna surgió la idea de que la amantadina pudiera ser de utilidad en los enfermos de Parkinson, puesto que a quienes se les prescribía***

para prevenir la influenza mostraban mejoría de sus manifestaciones neurológicas. Y b) *las observaciones de estudios poblacionales en los que se llega a establecer una asociación directa (o inversamente) proporcional entre la frecuencia de una enfermedad y alguna condición ambiental. Ha sido el caso de una relativa menor frecuencia de enfermedad coronaria en poblaciones con mayor ingesta de alimentos ricos en antioxidantes, o de cáncer de colon y una dieta con alto contenido en fibra.*

En cualquier caso, sea cual sea el origen de la hipótesis sobre el beneficio de una terapia, lo esencial es que estas deben probarse, y demostrarse como ciertas, mediante estudios clínicos, es decir, mediante la observación sistematizada y objetiva de su efecto en seres humanos que la reciben y su comparación con lo que habitualmente sucede en un grupo de enfermos sin la terapia.

Hay circunstancias en las que el beneficio de una cierta terapia es tan evidente e incuestionable que resulta innecesario realizar un ensayo clínico para aceptar su uso generalizado, tal como ha sucedido con el beneficio de los antibióticos en el tratamiento de la neumonía bacteriana, o de la cirugía abdominal en la apendicitis. Sin embargo, esto es algo excepcional, ya que lo normal es que la enfermedad responda de un modo distinto e impredecible, de modo que esta incertidumbre sobre el pronóstico no nos va a permitir saber qué grado de respuesta se debe a

la terapia de estudio, o a la propia evolución natural de la enfermedad, o consecuencia de otros determinantes ajenos a la intervención de la evaluación. Así, un grupo de enfermos puede evidenciar una mejoría en la respuesta clínica, bien por estar bajo el efecto placebo de la nueva terapia, o por estar en una fase muy precoz de la enfermedad, o porque reciben medicación concomitante. Es en estos casos en los que resulta indispensable –ética y científicamente– llevar a cabo el diseño, la conducción y la publicación formal de un estudio clínico experimental. El ensayo clínico controlado por sorteo es el diseño metodológico más confiable para distinguir si el beneficio atribuible a un tratamiento es real o solo un espejismo.

- **Estudios experimentales en la comunidad:** se trata de un tipo de estudios de asentamiento comunitario cuyo principal objetivo es generar un cambio a nivel de la población general. Antes de hacer un estudio comunitario, debemos partir de estudios preliminares de laboratorio, o de campo hechos con voluntarios, que garanticen de alguna manera la inocuidad de la manipulación que se pretende. En este tipo de diseños, una o varias comunidades recibirán la intervención, mientras que otras servirán como control y se observará el desarrollo natural del proceso que se investiga. La comunidad de referencia, que debe ser similar en todos los aspectos a la comunidad de intervención o experimental, ha de estar separada de esta última para

evitar contaminaciones por las maniobras realizadas. En caso de existir aleatorización, son las comunidades enteras y no los sujetos las que se distribuyen en el grupo experimental y de control. Esto no garantiza en absoluto que la distribución de todos los factores que puedan influir negativamente en el estudio sea similar entre los grupos comunitarios, ya que en realidad depende de la similitud entre las colectividades seleccionadas por el investigador. El investigador no tendrá capacidad de decidir el tratamiento que ha de estudiarse en una comunidad u otra.

La comunidad propuesta para participar como grupo de intervención debe ser representativa de la población en la que se aplicarán los resultados obtenidos. Se ha prestar especial atención a los recursos de la comunidad para la consecución de un efecto determinado. La aplicación de una intervención que ha resultado satisfactoria en una comunidad de recursos inferiores a la media será igualmente factible en otras zonas. Además, las experiencias deberían llevarse a cabo en zonas con un riesgo elevado de enfermar y, en su caso, durante el periodo de incidencia máxima de la enfermedad. La comunidad debe ser informada de los objetivos de la experiencia por razones de ética y porque se conseguirá con ello una mayor cooperación en la investigación.

En los estudios de este tipo no es posible determinar el número de sujetos y el grado en que reciben la influen-

cia del factor específico. La evaluación de los resultados incluye la comparación de los indicadores que midan el efecto previsto para el factor de interés en la comunidad estudiada con los correspondientes a la otra comunidad que sirve como control.

2.2.5. ¿Tiene el estudio un diseño adecuado a su objetivo?

Tal como acabamos de ver en el apartado anterior, la investigación clínica debe comprender los conceptos habituales en el diseño de una investigación, ya se trate de ensayos comparativos aleatorizados u otros tipos de estudios clínicos, como los estudios de observación.

Una vez definido el tipo de estudio que se va a realizar y establecer las hipótesis de investigación, el investigador debe concebir la manera práctica y concreta de responder a las preguntas de investigación. Esto implica seleccionar o desarrollar un diseño de investigación y aplicarlo al contexto particular de su estudio. Diseño se refiere al plan o estrategia concebidos para responder a las preguntas de investigación[79]. El diseño señala al investigador lo que debe hacer para alcanzar sus objetivos de estudio, contestar los interrogantes que se ha planteado y analizar la certeza de las hipótesis formuladas en un contexto en particular.

El tipo de diseño elegido para un ensayo clínico determinado dependerá de numerosos factores, tales como la clase de

patología o intervención médica que se va a estudiar, tamaño de la muestra requerido, etc. El diseño del ensayo clínico establece cómo asignar los pacientes a los grupos de tratamiento y las tácticas estadísticas para analizar los datos, dividiéndose en estudios **comparativos**, **cruzados** y **por pares**[80]. Las características de cada uno de estos tres tipos han sido resumidas en la **tabla 13**.

La esencia de una investigación es tener una buena pregunta que contestar y haber elegido el diseño adecuado para hacerlo.

Entre los estudios intervencionistas, el ensayo clínico comparativo aleatorizado se considera el de referencia para responder a las preguntas de tratamiento. Los ensayos clínicos comparativos aleatorizados no son absolutamente necesarios para demostrar una concordancia científica significativa, pero se considera que son los más convincentes y dignos de crédito. Un solo ensayo clínico comparativo, amplio y bien realizado podría aportar suficientes pruebas para dar por sentada una relación entre la sustancia y la enfermedad, siempre y cuando vaya respaldado por un conjunto de datos procedentes de estudios de observación o mecanicistas. En este sentido, los estudios de cohortes y los de casos y controles se utilizan como paso previo a la elaboración de una hipótesis causal y un ensayo clínico, o para responder a preguntas clínicas de pronóstico o de etiología/daño.

Tabla 13. Características principales de los estudios comparativos, cruzados y por pares (modificada de Idoate e Idoipe [2002][80])

Tipo de estudio	Ventajas	Inconvenientes
Estudios comparativos		
Un grupo recibe el nuevo tratamiento y el otro grupo de pacientes recibe el tratamiento estándar o placebo	Resultados extrapolables a la población No requiere orden de llegada Permite aleatorización (si el número de pacientes es elevado) o estratificación (si es más bajo)	Requiere un número elevado de pacientes Es costoso
Estudios cruzados		
Within patients Cada sujeto recibe dos tratamientos diferentes uno después del otro En dos (o varias) etapas Cada paciente recibe más de un tratamiento, el orden de asignación es aleatorio	No influyen factores relevantes tales como criterios de inclusión y exclusión, ya que cada paciente sirve de su propio control Se reduce la variabilidad, por lo cual se necesita un tamaño inferior de la muestra para poder detectar diferencia estadística entre el grupo de intervención y el grupo control	Solo sirven para valorar tratamientos paliativos de conductas crónicas

Tabla 13. Características principales de los estudios comparativos, cruzados y por pares (modificada de Idoate e Idoipe [2002][80]) *(continuación)*

Tipo de estudio	Ventajas	Inconvenientes
Estudios por pares *(matched-pairs)*		
Agrupar a los sujetos con factores relevantes iguales por parejas (p. ej., dos mujeres operadas de cáncer de útero y mayores de 50 años). Cada sujeto de la pareja se asigna aleatoriamente a un grupo (tratamiento activo o control/placebo)	Los grupos son exactamente comparables con respecto a los factores relevantes. Cuando se analizan los resultados se pueden sacar algunas conclusiones, no solo sobre qué tratamiento es el mejor, sino también sobre el tipo de sujetos que responden mejor al tratamiento activo (p. ej., operadas o mayores de 50 años)	Inflexibilidad del diseño al requerir pares de sujetos con factores relevantes idénticos. Estos pares podrían ser encontrados fácilmente en un gran grupo de pacientes con una condición crónica determinada, por ejemplo, con una clínica diabética, pero cuando la condición es menos común puede ser imposible encontrar el suficiente número de parejas comparables

Es importante, por ello, identificar el tipo de diseño elegido por los autores del trabajo y comprobar que es el adecuado para tema planteado por el estudio. Esta identificación nos va a permitir llevar a cabo una evaluación crítica del tipo de conclusiones que se pretenden publicar. Así, si el estudio tiene un diseño de caso control, las conclusiones pueden llevar a la asociación estadística de riesgo, pero no a conclusiones de causalidad. Si el diseño corresponde al de un estudio prospectivo, controlado, aleatorizado y doble ciego deberemos seguir las instrucciones de la lista de comprobación CONSORT para así comprobar que el diseño contenga todos los elementos necesarios para que el estudio se lleve a cabo (v. apartado 3.6). Si nos preguntamos sobre una nueva técnica de diagnóstico, hemos de encontrar un estudio de la validez de dicha prueba frente a un patrón oro de diagnóstico.

Aunque el análisis de estos diseños puede parecer complicado, como lectores, no debemos asustarnos, ya que mucho de lo que conocemos hoy día como 'evaluación crítica' no es más que sentido común Además, encontrarás un análisis minucioso sobre los criterios de validez específicos para cada tipo de estudio en el capítulo 3 de este libro.

El tipo de intervención del estudio debe estar bien definido; hemos de saber con qué se compara, lo que incluye una completa descripción del tratamiento y, si fuera pertinente, del placebo. Para ello, el análisis del artículo ha de continuar con dos preguntas fundamentales:

- **¿Cuál es la intervención o maniobra que se está comparando en el estudio? y ¿contra qué intervención se está comparando?**

- **¿Qué resultado se midió y cómo se hizo dicha medición?**

Sobre la intervención o maniobra ejecutada

Al leer un artículo, todos tenemos la tentación de creer lo publicado como algo que verdaderamente ha sucedido. Sin embargo, no es infrecuente que los autores –consciente o inconscientemente– hagan un "arreglo cosmético" de sus decisiones y manejo, exagerando la originalidad, rigurosidad y potencial importancia del estudio. Tenemos varias opciones para evitar o minimizar esto:

- **Evaluar el rigor del método de obtención de los datos. Por ejemplo, los análisis de fichas no son fiables 100%.**

- **Evaluar si es suficiente la cantidad de información sobre la intervención efectuada –dosis del medicamento, naturaleza del placebo, características detalladas de la**

intervención–, y de los sujetos reclutados –número de pacientes a los que se les propone el estudio y los que finalmente participan, equilibrio por raza, sexo, etc.–.

- Evaluar si se controlaron los factores de confusión, por ejemplo si los pacientes del grupo de estudio y de control difieren solo en la intervención estudiada y no en otra intervención adicional.

- Evaluar la originalidad del estudio, descartando la existencia de otros similares.

También es fundamental que se describa claramente la variable principal de resultado y cómo se ha medido. En el apartado de las variables de evaluación se describe qué se va a registrar, cómo, cuándo, con qué y por qué, así como los detalles y especificaciones de la instrumentalización y del equipo necesario[81]. Esto se aplica tanto en el contexto de los estudios observacionales como de las variables en respuesta a una intervención. Y se refiere al surgimiento de los efectos y acontecimientos observables y mesurables, que representan las variables dependientes de la investigación. Lo recomendable es identificar una o dos primarias (variables de evaluación) y unas pocas secundarias (variables intermedias), tantas como objetivos fueron propuestos. Es pertinente elegir las más sensibles y estandarizadas en cuanto a los métodos de medición y proceder en la forma más estricta y rigurosa para su evaluación objetiva cuando llegue el momento de reunir datos. Así,

cualquier variable de evaluación ha de cumplir una serie de características ideales:

- Debe derivarse directamente del objetivo. Como principio es preferible que el objetivo del estudio se exprese en una sola variable de evaluación.
- Calidad de medida. La variable de evaluación ha de aportar sensibilidad, exactitud, consistencia, fiabilidad –reproducibilidad y repetitividad– y una baja variabilidad intraobservador e interobservador.
- No ha de ser invasiva y sí de fácil medición.
- Clínicamente relevante. La variable relevante es aquella que se corresponde con el objetivo y es la que interesa al paciente.

Respecto a las variables intermedias/subrogadas, las características que hemos de encontrar son las siguientes:

- Debe estar en algún paso fisiopatológico del desarrollo de la enfermedad.
- Debe ser un marcador directo del pronóstico de la enfermedad. Su magnitud en cualquier tiempo de la evolución de la enfermedad debe ser pronóstico del *outcome*, y la magnitud de la variación a lo largo del tiempo es proporcional al riesgo del *outcome*. Debe haber un efecto dosis-respuesta.

- La reducción/mejoría en la variable intermedia, en respuesta al tratamiento debe conllevar o tener una relación directa con la probabilidad del *outcome*.
- Deberíamos conocer adecuadamente esa relación.
- Debe ser posible su utilización en la práctica clínica.

Algunos ejemplos de variables clínicas e intermedias han sido expuestos en la **tabla 14**.

La variable de resultado ha de ser importante desde el punto de vista clínico y para el paciente debe ser objetiva; cualquier cambio que experimente ha de reflejar cambios reales en el paciente. El uso de variables poco importantes para el paciente o subrogadas de escaso impacto clínico puede hacer que perdamos el interés por leer el artículo.

Los métodos de medida tienen que ser descritos con detalle y deben ser evaluados críticamente en busca de errores que afecten a la validez y fiabilidad de los mismos. Una variable de medida válida debe medir aquello que se supone que tiene que medir. Una variable de medida fiable debe dar un resultado similar tras varias repeticiones. Debemos prestar especial atención a aquellas variables de medida dependientes de

valoraciones subjetivas del examinador por estar sujetas a mayor riesgo de error.

Tabla 14. Tipos de variables clínicas e intermedias

Enfermedad	Variable intermedia	Variable clínica
Arritmias	Arritmias ventriculares	Supervivencia
Insuficiencia cardiaca crónica	Fracción de eyección del ventrículo izquierdo	Supervivencia
Hipertensión arterial	Tensión arterial	Infarto agudo de miocardio Supervivencia
Cáncer de próstata	Biopsia de próstata	Supervivencia
Virus de la inmunodeficiencia humana	Células CD4	Infecciones oportunistas Supervivencia
Asma infantil	Espirometría forzada Prueba broncodilatadora	Sibilancias durante los últimos 12 meses
Dermatitis atópica	Concentración de IgE	Prurito en los últimos 12 meses

Sobre la medición del resultado

Los datos generados durante la fase experimental o de campo requieren de un procesamiento y análisis, por lo que deberá especificarse dónde y cómo se procesarán esos datos para así obtener los resultados.

Una vez concluidas las etapas de recolección y procesamiento de los resultados se inicia con una de las más importantes fases de una investigación: el análisis de datos, que es el precedente para la actividad de interpretación.

La interpretación se realiza en términos de los resultados de la investigación. Esta actividad consiste en establecer inferencias sobre las relaciones entre las variables estudiadas para extraer conclusiones y recomendaciones[82]. La interpretación se realiza en dos etapas: *a)* interpretación de las relaciones entre las variables y los datos que las sustentan con fundamento en algún nivel de significación estadística, y *b)* establecer un significado más amplio de la investigación, es decir, determinar el grado de generalización de los resultados de la investigación.

Es decir, en concordancia con el tipo de variable aplicada y al amparo de los objetivos del estudio, se propone a priori una breve exposición del análisis estadístico por aplicar. En gene-

ral, se aplicaría la estadística descriptiva como una primera aproximación para valorar el comportamiento general y los resultados obtenidos con cada variable. Después se procedería con la estadística inferencial, a partir de la aplicación de pruebas estadísticas paramétricas o no paramétricas, según el tipo de dato disponible y las características de cada variable por analizar. También, en este momento deben aparecer definidos los criterios para la toma de decisión estadística (nivel de significación o valor alfa).

Debe evaluarse si el resultado que se ha medido es importante de acuerdo con el problema de investigación que suscita el estudio. Así, si lo que se prueba es un tratamiento nuevo para combatir una enfermedad mortal, el resultado clínicamente relevante que evaluar es el tiempo de sobrevida y la calidad de vida. Por el contrario, medir la concentración de una proteína plasmática que potencialmente se correlacione con la posibilidad de una mayor sobrevida no va a resultar interesante.

La medición de un efecto sintomático (prurito), funcional (movilidad), psicológico (ansiedad) o social (inadaptabilidad social) resulta ciertamente difícil. En estos casos, el artículo debe aportar datos de evidencia de que el resultado proceda de una herramienta previamente validada. Por ejemplo, cuando se trata de medir el asma, necesitaremos una herramienta que nos sirva de guía para el tratamiento y facilite el control de la enfermedad. En este sentido, el estudio español COVA-

LAIR[83], cuyo objetivo principal consistió en evaluar las propiedades de medición de las versiones en español del *Asthma Control Test* para su utilización en la práctica clínica habitual, concluyó que se trata de una escala validada para medir el control del asma. Permite discriminar adecuadamente a los pacientes mal controlados, es fácil de aplicar a los enfermos y se correlaciona bien con la gravedad, los síntomas y algo fundamental en medicina, con la valoración del profesional médico y del paciente. Para esta y cualquier otra escala, hemos de confirmar que mide lo que realmente deseamos que mida, que cambios en la medida del resultado reflejen verdaderos cambios en el estado del paciente.

Además, el análisis de los resultados debe hacerse conforme al protocolo inicial. En palabras de Molina Arias, en su artículo *Lectura crítica en pequeñas dosis: Razones para dejar de leer un artículo*[84] decía: "... debemos dudar siempre de las conclusiones de análisis parciales no especificados antes de iniciar el ensayo... no es correcto torturar los resultados hasta que confiesen...".

Para terminar con la valoración crítica del diseño del estudio, debemos comprobar que los autores cumplen con el protocolo definido previamente al inicio del ensayo. Cualquier violación del protocolo ha de estar justificada, de modo que cualquier cambio en las condiciones del ensayo, bien por las dificultades encontradas durante su desarrollo, o bien por el interés de los propios autores, debe hacernos desconfiar

de los resultados obtenidos. Parece razonable interrumpir un ensayo si se registraran efectos adversos graves para los participantes del estudio, pero es más difícil justificar la interrupción precoz por encontrar los resultados buscados por los autores.

El apartado de acontecimientos adversos se requiere en estudios experimentales y se contextualiza como riesgos derivados del estudio. Aplica a cualquier experiencia no deseable o involuntaria que le ocurre al sujeto en el curso del estudio (incluye hallazgos de laboratorio anormales), sobre todo cuando se trata de experimentar un medicamento o un procedimiento. Además del registro de las variables, es necesario anticipar el tiempo y espacio para describir detalladamente sus características, la duración, intensidad, frecuencia y la gravedad, junto con los parámetros para analizar la relación de causalidad (efecto adverso), pues todos estos datos deberán consignarse de forma individualizada, y además, notificarse a la autoridad sanitaria cuando son eventos graves e inesperados. Estos sistemas de notificación sirven para compartir información y resolver problemas que afectan a la seguridad de los pacientes. La notificación es obligatoria cuando se trata de reportar eventos adversos graves –infecciones nosocomiales, amputación, muerte neonatal posparto, reacción adversa a un fármaco, daño cerebral tras una cirugía–, mientras que la notificación

de incidentes menos graves, que producen daño leve o temporal –caídas de pacientes, aplicación del medicamento equivocado, identificación equivocada de paciente– es voluntaria.

Cabe aclarar que el apartado que describe los tratamientos y, por tanto, el registro sistemático de eventos adversos, no se aplica a los estudios observacionales. En el marco de una investigación observacional se recoge información sobre el curso de la evolución o de los medicamentos que el paciente ya está recibiendo por su condición clínica, y que han sido prescritos en el contexto de la práctica médica habitual por su seguridad y eficacia ya probadas.

El seguimiento ha de ser entonces lo más completo posible, de modo que los autores han de dar cuenta del abandono por parte de los participantes del estudio, ya sea por no cumplimiento del tratamiento o de cualquiera de las condiciones de planteamiento y ejecución del protocolo. Es frecuente tener pérdidas a lo largo de un estudio, que deben ser reflejadas preferentemente en organigramas. Es recomendable revisar las posibles discrepancias respecto al total de sujetos evaluados, que pueden indicar descuidos de los autores y/o errores de impresión y pueden afectar a la validez de las conclusiones.

Sin duda, la etapa más crítica de un estudio corresponde a la del diseño, ya que resulta casi imposible corregir a posteriori los sesgos introducidos durante esta etapa. Veamos en qué consiste eso de los sesgos de un trabajo de investigación en el siguiente apartado.

2.2.6. ¿Se han evitado o minimizado los sesgos?

O lo que es lo mismo: ¿Se han incluido estrategias para optimizar la validez y precisión del estudio? Diariamente, la ciencia rinde un amplio caudal de evidencia cuya calidad suele ser heterogénea. Ahora bien, ¿a qué nos referimos cuando hablamos de la calidad de la evidencia? Podríamos considerar que es la serie de atributos que nos permiten confiar, en menor o mayor grado, en los resultados obtenidos en los diferentes estudios. Por esta razón, diseñar un estudio implica planificar la manera en que los diferentes sesgos serán evitados y cómo podrán interpretarse los resultados a pesar de los sesgos que no pudieron ser eliminados.

Los individuos somos y reaccionamos de un modo diferente y, por ello, la medicina no es una ciencia exacta. Para describir con un solo dato qué ocurre en un grupo de individuos, calculamos porcentajes de individuos que han sido vacuna-

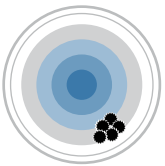
dos, medias de tensión arterial, porcentajes de pacientes que cumplen un tratamiento y se curan, etc. Para la estimación de estas medias y porcentajes se selecciona una muestra de la población global, lo más representativa posible, pero nada garantiza que esta muestra sea efectivamente representativa. Solo por azar existe una posibilidad de error y, por ello, se llama **error aleatorio**. Este tipo de error se cuantifica en el análisis del valor de la ***p*** o la estimación del intervalo de confianza, llamado así porque es un rango de valores en el que confiamos que se encuentre el verdadero valor. Cuanto mayor sea el valor de ***p***, mayor es la probabilidad de que los resultados obtenidos estén afectados por el error aleatorio. Por consenso se acepta que $p < 0{,}05$ es estadísticamente significativo, es decir, refleja a la población de la que fue obtenida la muestra.

Cuanto mayor es el tamaño de la muestra estudiada, menor es el error aleatorio, más estrecho es el intervalo de confianza, mayor es la precisión en la estimación del dato, hasta el punto de que si lográsemos estudiar a toda la población no existiría la posibilidad de equivocación; el error aleatorio sería nulo.

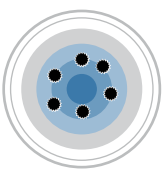
Pero con estudiar grandes muestras no se soluciona el problema; también existe la posibilidad de cometer lo que se denominan **errores sistemáticos o sesgos**. El sesgo *(systematic bias)* es un error sistemático que nada tiene que ver con el tamaño de la población. Se relacionan con el diseño, desarrollo o análisis del estudio. Como conllevan una estimación incorrecta del riesgo asociado a una exposición, hacen que los resultados del estudio se desvíen o se aparten de la realidad. En función de su importancia condicionan las posibles limitaciones a la validez interna de los estudios y habitualmente determinan también una baja validez externa[85]. Los sesgos suelen producirse como resultado de errores en diferentes aspectos del diseño: estructura del estudio, selección y valoración de los pacientes, método de recogida de los datos pertinentes a la exposición o enfermedad y análisis de los datos. En consecuencia, los resultados observados del estudio tenderán a ser diferentes de los resultados verdaderos **(figura 9)**.

El tiro al blanco y los errores de medición

Error sistemático



Error aleatorio



Estos dos tipos de errores, aleatorio y sistemático, no solo afectan al muestreo, sino también a la medición. El **error aleatorio** ocurre cuando las mediciones repetidas, ya sean en un mismo sujeto o en diferentes miembros de la población en estudio, varían de manera no predecible, mientras que el **error sistemático** ocurre cuando estas medidas varían de manera predecible y, por lo tanto, se tiende a sobreestimar o subestimar el valor verdadero en medidas repetidas. El **"tiro al blanco"** es un buen ejemplo para explicar ambos conceptos, en el que el punto medio de la diana (objetivo) es el valor verdadero en la población blanco y los "disparos" son las mediciones tomadas de la población de estudio para estimar dicho valor verdadero. Un buen tirador, pero con un arma mal calibrada, podrá ser muy preciso (todos los disparos van al mismo punto), pero ninguno de ellos dará en el blanco correcto. Esto corresponde al **ERROR SISTEMÁTICO**. Por el contrario, un tirador poco experimentado, pero con un arma bien calibrada, estará apuntando al blanco correcto aun cuando sus disparos no den en el punto medio del blanco seleccionado. Esto corresponde al **ERROR ALEATORIO**.

Figura 9. El tiro al blanco y los errores verdaderos.

Cuando interpretamos los resultados de un estudio nos podemos preguntar: ¿Podrían deberse los resultados a algo que los autores no han tenido en consideración?, como por ejemplo:

1. Los grupos del estudio no son comparables debido a cómo fueron seleccionados los pacientes (sesgos en la selección).

2. Los grupos de pacientes del estudio no son comparables debido a cómo se obtuvieron los datos (sesgos en la información).

3. Los autores no han recogido información (o la han obtenido pero no la han utilizado) sobre un factor que se relaciona a la vez con la exposición y con el efecto estudiados (factor de confusión).

- **Sesgos de selección:** es un error sistemático en el proceso de selección de los individuos que limita la comparación de los grupos. La muestra seleccionada (aun siendo de tamaño más que suficiente) no es representativa de la población de la que se quieren extraer conclusiones. Es un sesgo estadístico en el que hay un error en la elección de los individuos o grupos que van a participar en un estudio científico[86]. A veces se le conoce como efecto de selección. Casos típicos en los que se puede dar este tipo

de sesgos son: casos hospitalarios, exclusiones debidas a fallecimiento del paciente antes de su llegada al hospital, exclusión por no estar lo suficientemente enfermo como para ingresar en el hospital, exclusiones debidas a la distancia, coste u otros factores.

A la hora de preparar una publicación, existe un tipo especial de sesgo de selección que consiste en la inclusión selectiva de los trabajos basados en criterios subjetivos del revisor (sobre la calidad de los estudios), lo que origina diferencias subjetivas en el peso de los estudios. La muestra puede, independientemente de los factores externos al grupo de revisores, sesgarse a favor de los auspiciadores de la investigación y en ese caso, la falta de ética científica, es manifiesta.

Veamos el siguiente ejemplo verificado en la literatura que ilustra la incidencia de dicho sesgo: el bioquímico Linus Pauling –premio Nobel de Química en 1954 por su investigación en la naturaleza del enlace químico, y Premio Nobel de la Paz ocho años más tarde por su campaña contra las pruebas nucleares terrestres– usó citas selectivas de la literatura médica para "demostrar" su teoría de que la vitamina C ayuda a sentirse mejor y vivir más tiempo (mayor bienestar, más salud). Knipschild et al., citados en Greenhalgh[86], llevaron a cabo una revisión sistemática en la literatura para encontrar evidencias y contrastar los resultados alcanzados por Pauling. Encontraron que, aunque

uno o dos ensayos sugirieron fuertemente que la vitamina C podría prevenir el ataque de catarro-gripe, había más estudios que no mostraron efecto beneficioso.

Los sesgos de selección pueden ocurrir en cualquier estudio epidemiológico. Sin embargo, ocurren con mayor frecuencia en estudios retrospectivos y, en particular, en estudios transversales o de encuesta[87]. En los estudios de cohorte prospectivos los sesgos de selección ocurren raramente, ya que el reclutamiento y selección de la población en estudio se da antes de que ocurra el evento en estudio. En los estudios retrospectivos los sesgos de selección pueden ocurrir cuando los participantes potenciales o los investigadores conocen la condición de exposición y/o de enfermedad, y este conocimiento influye diferencialmente en la participación en el estudio. Los estudios de casos y controles son particularmente susceptibles de este tipo de sesgo ya que en la mayoría de sus aplicaciones se trata de estudios retrospectivos. En el contexto de este tipo de estudios, el sesgo se presenta cuando el procedimiento utilizado para identificar el estado de enfermedad (sesgo diagnóstico) varía o se modifica con el estado exposición. Este sesgo se llama *sesgo de detección*.

Tanto la falta de respuesta por parte de los participantes de un estudio como la detección diferencial, originada cuando la prueba diagnóstica para detectar el evento se realiza con mayor frecuencia en el grupo expuesto, pueden también introducir a un sesgo de selección. Si el sesgo de selección no

se tiene en cuenta, todas las conclusiones que salgan de un estudio pueden estar mal[88].

- **Sesgo de información:** Es un error sistemático en la medición de la variable. Todas las variables (predictoras y de resultado) deben medirse de la misma forma en todos los participantes del estudio. Esto conduce a una medida de la variable de estudio que es diferente de su valor verdadero y puede producir una clasificación errónea de los sujetos de estudio con respecto a la exposición, la enfermedad o ambos[88]. Ocasionan una conclusión errónea respecto de la hipótesis que se investiga. Una posible fuente de sesgo de medición puede ser cualquier factor que influya de manera diferencial sobre la calidad de las mediciones que se realizan en los grupos expuesto y no expuesto en el contexto de los estudios de cohorte o entre los casos y controles en el contexto de los estudios de casos y controles. Aunque prácticamente no existen procedimientos libres de error de medición, no todos los errores de medición son fuente de sesgo de información. Los errores de medición pueden ser no diferenciales (aleatorios) cuando el grado de error del instrumento o técnica empleada es el mismo para los grupos que se comparan; y diferenciales (no aleatorios) cuando el grado de error es diferente para los grupos estudiados. El sesgo de información se refiere particularmente a este último tipo[86].

El sesgo de información conduce a errores de la clasificación y/o desenlace. La magnitud de la clasificación errónea está determinada por la sensibilidad y la especificidad del método usado para realizar dicha clasificación y aumenta cuando disminuye la prevalencia de exposición en los controles. El error de clasificación no diferencial ocurre cuando el grado de error de clasificación de la exposición es independiente de que se trate de un caso o un control. El error de clasificación diferencial ocurre cuando el grado de error de clasificación difiere entre los grupos que se comparan.

Una medición inexacta puede influir en la valoración de un desenlace. La información para medir un desenlace puede proceder de tres fuentes distintas: *a)* lectura de los instrumentos de medida; *b)* mediciones del investigador, y *c)* informes o registros obtenidos de individuos. La información obtenida puede ser inexacta, porque los datos producidos estén sistemáticamente fuera del objetivo y en la misma dirección debido a un sesgo por cómo se recogieron los datos.

El error de medición también se puede deber a la falta de exactitud de los instrumentos empleados. Defectos en la sensibilidad del instrumento de medida, o la falta de calibración, hacen que no se pueda detectar lo que se busca, o se haga sin exactitud ni precisión (p. ej., una balanza mal calibrada dará sistemáticamente un error de medida del peso de los sujetos del estudio). El uso de encuestas o cuestionarios no validados, así como de instrumentos inadecuados reportará

datos no exactos y no confiables. Valga como ejemplo que si utilizáramos la radiografía gastroesofágica para valorar los efectos secundarios de un antiinflamatorio, no sería válido, ya que el "instrumento" de diagnóstico no es el adecuado para evaluar la gastritis.

La posibilidad de un sesgo del investigador existe en todos aquellos casos en que la medida del desenlace depende de que el investigador interprete subjetivamente los datos. Sin embargo, es posible reconocer y corregir un principio fundamental de la psicología humana: el de que todos nosotros, investigadores incluidos, vemos aquello que queremos o esperamos ver. Esto se consigue evitando que el investigador que valora el desenlace conozca el grupo al que se ha asignado un individuo. La valoración a ciegas puede emplearse en los estudios de casos y controles y en los de cohortes, así como en los ensayos clínicos aleatorios. No utilizarla puede conducir a este tipo de sesgo.

La información dada por los sujetos del estudio está ligada a los *sesgos de recuerdo y de declaración*. El *sesgo de recuerdo* presupone defectos de memoria, en particular, cuando es más probable que los individuos de un grupo no recuerden ciertos sucesos que los de otros grupos sí recuerdan.

En un estudio llevado a cabo en madres de niños con espina bífida, los autores concluyeron que habían demostrado una asociación entre el dolor

de garganta durante el embarazo y tener un bebé con espina bífida. ¿Cómo fue posible llegar a semejante conclusión? De las madres de hijos con espina bífida, el 50% declararon haber padecido dolor de garganta durante el embarazo, mientras que entre las madres de hijos sin la enfermedad solo lo declararon 5%. Se puede afirmar que es más probable que las madres que experimentaron el traumatismo de tener un hijo con espina bífida reactivaran su memoria y recordaran sucesos que habitualmente no se recuerdan con más intensidad que el resto de mujeres.

El *sesgo de declaración* se produce en los estudios de casos y controles cuando los sujetos de un grupo de estudio relatan con más exactitud sus recuerdos que los del otro grupo. Aquí interviene un punto débil fundamental en los estudios publicados y es la fatiga, es decir, hay pacientes que ya han formado parte de otros estudios sin éxito, que se encuentran cansados de tanto seguimiento y solo dan parte de la información.

En un estudio sobre la relación entre la gonorrea y el número de compañeros sexuales, a las participantes del estudio se les informó de que solo se podían prevenir las graves consecuencias de la enfermedad si se localizaba y se trataba a sus compañeros sexuales. Las mujeres con gonorrea declararon haber tenido en promedio el doble de compañeros sexuales que las mujeres sin gonorrea. Es más probable que el sesgo de

declaración se cometa cuando la información que se busca es personal o delicada. Se ve cómo en este estudio las mujeres con gonorrea se sintieron más obligadas y, por tanto, menos reacias a ofrecer información acerca de sus compañeros sexuales que las mujeres sin gonorrea.

También puede ocurrir que individuos asignados inicialmente a un grupo particular deciden migrar de grupo por preferir un tipo de intervención sobre otro. A este se le identifica como *sesgo de adaptación*, y se produce especialmente en estudios de intervención (experimentales o cuasiexperimentales).

Cuando el seguimiento de los pacientes es incompleto, por ejemplo por fallecimiento, existe la posibilidad de que la frecuencia del desenlace en los que no fueron incluidos en la valoración final sea distinta de la frecuencia en los que fueron incluidos. El problema de una observación desigual de los dos grupos puede invalidar los resultados. Así, una frecuencia más alta de visitas al hospital hace que aumente la frecuencia de las observaciones, aumentando con ello el recuento de los síntomas más comunes –cefaleas, fatiga, depresión, etc.– que se miden en muchos estudios.

Aunque el resultado de un estudio cumpla los difíciles criterios de una valoración apropiada, exacta y completa, nos queda aún sitio para la preocupación. Los investigadores intentan medir los sucesos como habrían ocurrido si nadie los hubiese observado. Desafortunadamente, el proceso real de

llevar a cabo un estudio puede incluir la introducción de un observador en los sucesos que se miden. Por ello, antes de que un artículo caiga en nuestras manos, el revisor de dicho artículo ha de determinar si el proceso de observación modificó el resultado.

- **Sesgo o factor de confusión:** el fenómeno de la confusión implica la posibilidad de que la explicación del efecto encontrado se deba a otros factores. Existe sesgo de confusión cuando observamos una asociación no causal o falsa entre la exposición y el evento en estudio, por la acción de una tercera variable no controlada o *factor de confusión*. Los resultados de un estudio estarán confundidos cuando los resultados obtenidos en la población en estudio apoyan una conclusión falsa sobre la hipótesis en evaluación, debido a la influencia de otras variables que no fueron controladas adecuadamente ya sea durante la fase de diseño o de análisis. Todos los resultados derivados de estudios observacionales están potencialmente influidos por este tipo de sesgo. En los estudios observacionales el sesgo de confusión se puede entender como un problema de comparabilidad cuyo origen está ligado a la imposibilidad de realizar una asignación aleatoria de la exposición en los sujetos de estudio[87]. Por ejemplo, si se estudiara directamente la relación entre beber alcohol y el cáncer de pulmón se encontraría una asociación positiva. Es decir, aquellos pacientes que beben tienen un mayor riesgo

de presentar cáncer de pulmón. Sin embargo, cuando se ajusta según el consumo de tabaco este efecto desaparece. La explicación de este fenómeno sería que los bebedores de alcohol tienen una mayor probabilidad de fumar más y el fumar sí que tiene una relación etiológica con el cáncer de pulmón.

¿Pero qué es entonces una variable confusora o confusor? Se trata de variables que están asociadas con el episodio en estudio de manera causal, es decir, son factores de riesgo para el evento en estudio (fumar es un factor de riesgo del cáncer de pulmón); debe asociarse con la exposición en la población a estudio (el consumo de alcohol se asocia al consumo de tabaco) y, además, esta asociación debe ser por un mecanismo causal diferente del que actúa sobre la exposición en estudio y debe ser independiente de la misma (el tabaco no forma parte del supuesto mecanismo de acción del alcohol para aumentar el riesgo de cáncer de pulmón).

Los sesgos de confusión en estudios de cohorte ocurren principalmente cuando el grupo expuesto y el no expuesto no son comparables. Por ejemplo, si fuera de interés evaluar en un estudio de cohorte el efecto del consumo de café sobre el riesgo de infarto de miocardio, se tendría que considerar el tabaquismo un posible factor de confusión por estar asociado con el consumo de café y es de por sí un factor de riesgo para infarto de miocardio.

Otro ejemplo clásico es la vacunación contra tuberculosis que se realizó para el personal de cierto centro hospitalario. Una vez vacunadas las personas se encontró que un importante número desarrolló hepatitis B. En este caso, ¿era la vacuna contra la tuberculosis una causa de hepatitis o existía un evento o variable que estaba relacionado con ambas variables y que en verdad ocasionaba el efecto? En este caso, se observó que la inoculación del virus se dio a través de las jeringas que se usaron para vacunar al personal: este era el factor de confusión. Tras observar este ejemplo se pone de manifiesto el efecto negativo que tendría en la toma de las decisiones el no vacunar al personal contra la tuberculosis cuando el problema no era el vial de la vacuna sino el vehículo.

¿Cómo controlar los sesgos?

Los sesgos se pueden cometer en cualquier etapa de una investigación, por lo que la mejor manera de evitarlos es realizar un buen diseño de la investigación que se va a llevar a cabo. Debemos entender que la inmensa mayoría de los sesgos de selección y de información no pueden ser corregidos una vez han sido recogidos los datos. Por tanto, es imprescindible reflexionar sobre los posibles sesgos antes de empezar a recoger los datos y realizar un diseño adecuado que minimice la probabilidad de cometerlos.

Disponemos de distintas estrategias para controlar los sesgos de selección y de evaluación:

- **Aleatorización**. Es el método electivo ideal para alcanzar la comparabilidad entre los grupos. Fortalece las pruebas estadísticas empleadas, disminuye el sesgo de selección e intenta proteger contra otro tipo de sesgos, fundamentalmente el de confusión (su principal finalidad). Matemáticamente, se define como que todos los individuos tengan la misma probabilidad (p) de entrar en el grupo de tratamiento y al grupo control ($1 - p$). La aleatorización total es el procedimiento normalmente recomendado, pero presenta algunos problemas, como la falta de aceptación por parte de los sujetos al no saber a priori el tratamiento que van a recibir. Esto causa una reducción en el tamaño de muestra que puede ser importante. Además, la aleatorización total normalmente produce grupos desiguales en el tamaño de la muestra, tanto más cuanto más reducido es el número de participantes.

La aleatorización persigue como objetivos disminuir los sesgos de selección (rechazar pacientes por conocer a qué tratamiento será asignado), evitar sesgos en la asignación a los tratamientos y la homogeneidad entre los grupos en comparación.

- **Enmascaramiento.** La forma de controlar estos sesgos producto de la subjetividad de los individuos involucrados en la investigación es la técnica de enmascaramiento *(masking)*. Este procedimiento consiste en ocultar, al conocimiento del sujeto experimental, la identidad de las intervenciones que se van a ejecutar en el estudio *(ciego simple)*. El estudio ciego simple intenta evitar la influencia del conocimiento del grupo en la respuesta del participante. Si, además, esta identidad es desconocida por el individuo que administra la intervención (investigador) estamos en presencia del *doble ciego*, que trata de eliminar la tendencia a recoger los datos de manera diferencial, favoreciendo la hipótesis de trabajo. Finalmente, si le agregamos el desconocimiento de la persona que procesa los datos (operador), estaremos frente a un experimento con técnica de enmascaramiento triple ciego. En el enmascaramiento *triple ciego* el investigador que analiza los resultados no sabe cuál es, entre los datos que se le ofrecen, el que mide el efecto que se analiza.

Tras haber visto las distintas estrategias de control sobre los sesgos de selección y evaluación, cabe preguntarnos entonces qué podemos hacer para prevenir la ocurrencia de sesgos de confusión. Las variables de confusión pueden controlarse tanto en la fase de diseño de la investigación –aleatorización, restricción y pareamiento– como en la fase de análisis –estandarización, estratificación y modelos de análisis multivariante–.

Incluso los intentos más acuciosos por generar grupos comparables serán perdidos, especialmente si la persona que mide el resultado sabe a qué grupo pertenece el paciente. El investigador podría reevaluar el resultado si este es inapropiado al tratamiento que se está evaluando. Si, por ejemplo, se está probando una droga hipotensora, y la paciente es del grupo estudio, pero la presión arterial está más alta, el evaluador tenderá a medir nuevamente la presión.

Es importante que se haga el máximo esfuerzo posible por lograr el "ciego" tanto al paciente, como al médico, e idealmente al analista de los datos. Cuando hacemos lectura crítica de un artículo siempre hemos de tener en mente que "no es correcto torturar los resultados del estudio hasta que confiesen".

Respecto al control en la fase de diseño:

- **Aleatorización.** Aquí la aleatorización intenta conseguir que todos los factores de confusión se distribuyan por igual en los grupos de estudio. Si el tamaño de la muestra es lo suficientemente grande, un experimento bien diseñado puede esperar controlar todos los efectos de

distorsión producidos por factores de riesgo extraños al estudio (sean o no de confusión), incluidos aquellos que no sean medidos. El control de factores de confusión por la aleatorización es directamente proporcional por tanto al tamaño de la muestra.

- **Restricción.** Consiste en seleccionar para el estudio la población que está en una única categoría de la variable de confusión. Por ejemplo, si fuera de interés evaluar en un estudio de cohorte el efecto del consumo de café sobre el riesgo de infarto del miocardio y se establecieran dos grupos de contraste, uno conformado por bebedores de alcohol y otro por no bebedores de alcohol, el tabaquismo se tendría que considerar un posible factor de confusión, ya que está asociado con el consumo de alcohol y es un factor de riesgo para infarto de miocardio.

- **Apareamiento.** En el apareamiento se comparan dos grupos de población que se diferencian en la exposición al factor de riesgo. Iguala en el proceso de selección a los grupos de comparación con relación a los factores de confusión. Volviendo al ejemplo anterior: un grupo consume alcohol y otro no, pero son idénticos en cuanto a la variable de confusión (selección únicamente de sujetos fumadores).

Respecto al control en la fase de análisis:

- **Estandarización.** También conocida como ajuste, la estandarización nos permite comparar los grupos de estudio si la distribución de la variable de confusión fuera la misma en ambos. Permite obtener medidas del efecto una vez eliminada la influencia de la variable de confusión (medidas ajustadas). Por ejemplo, la mortalidad en una población está condicionada por su edad. Para eliminar la influencia de la edad, es necesario comparar tasas de mortalidad estandarizadas, es decir, comparar tasas calculadas en el supuesto de que ambas poblaciones de estudio tuvieran la misma distribución de edad.

- **Estratificación.** Estima la medida del efecto en subgrupos de estudio que son similares con relación a los factores de confusión. En la fase de análisis podemos realizar este tantas veces como categorías tenga la variable de confusión. Por ejemplo si la variable de confusión es el sexo (variable con dos categorías), el análisis se puede hacer separadamente en hombres y mujeres. Si se busca la asociación entre infección por el hongo *Aspergillus* y asma alérgica hacemos el análisis para hombres: comparamos en la muestra de hombres un grupo de infectados con otro grupo de no infectados y vemos la aparición de asma, e idénticamente para las mujeres infectadas y no infectadas, de forma independiente.

- **Análisis multivariante.** Permite determinar la influencia sobre el efecto de más de un factor de riesgo. Permite estimar parámetros ajustados para cada una de ellas a igualdad de situación respecto a las otras variables de confusión.

Este tipo de análisis es especialmente útil en algunas áreas de la medicina como la alergología, por permitir la inclusión de variables que otros análisis previos hayan identificado como significativamente diferentes entre los grupos de estudio y también factores que son considerados asociados con el desarrollo de alergias. Este es el caso de un estudio de casos y controles que analice la relación entre lactancia materna, alimentación complementaria y desarrollo de alergia.

2.2.7. ¿Es adecuada la población estudiada?

La pregunta de investigación que se ha planteado en el artículo que estamos leyendo es el núcleo central de la investigación, nos indica los posibles diseños de investigación que podemos seguir y nos delimita cuál es nuestra población de estudio.

Antes de determinar si el resultado se aplica a su población y a su práctica clínica, debemos valorar cuáles fueron los pa-

cientes incorporados o los potencialmente incorporados al estudio. ¿Pero con qué criterio podemos evaluar críticamente si la población que representa el estudio es correcta o incorrecta?

Para hacer este análisis deberíamos responder a distintas preguntas:

- **¿Cómo fueron elegidos los pacientes?**
- **¿Quién fue incluido/excluido en el/del estudio?**
- **¿Cómo se hizo el muestreo?**
- **¿Es adecuado el tamaño de la muestra?**

¿Cómo fueron reclutados los pacientes?

Para responder a esta pregunta, hemos de tener claro primero cómo definir la *población de un estudio*. Podemos definirla como un grupo de casos que cumple con una serie predeterminada de criterios. A la población del estudio se le conoce como *población blanco o diana*, y es el universo al que queremos generalizar nuestros resultados. Dentro de esta población diana se encuentra la *población accesible*, que consta del conjunto de casos que satisfacen los criterios predeterminados y que, al mismo tiempo, son accesibles para el investigador. Esto es importante porque el modo de selección de

los pacientes determina si hubo o no un sesgo de selección (véase el apartado anterior). Interesa que no haya un sesgo de selección, de manera que los resultados del estudio no estén motivados por las características especiales de la población reclutada.

En un grupo en el que la población diana sean niños con alergia al polen, la población accesible pueden ser niños alérgicos que acudan a los controles médicos pertinentes en un área geográfica determinada.

¿Quién fue incluido/excluido en el/del estudio?

Al identificar una población, el investigador debe especificar bien los criterios que definen quiénes deben incluirse, ya que mediante ellos se decidirá si un individuo se incluirá o no en la población de interés. Estos criterios de selección han delimitado a la *población elegible*. Los *criterios de elegibilidad o de inclusión* son los que definen las características que debe tener la población (p. ej., tener más de 18 años, padecer una enfermedad cardiovascular, haber sido diagnosticado de dermatitis atópica, etc.). Aunque a veces una población se elige por las características que sus miembros no deben tener, y para ello se establecen los *criterios de exclusión* (p. ej., en nuestro país personas que no hablen español, tener más de

70 años, haber tomado una medicación reciente que pueda interferir con la variable que se busca, etc.).

Si los criterios de inclusión son muy estrictos, la población elegible no será representativa de todos los pacientes, y los resultados solo serán aplicables a ese grupo. Si evaluáramos la farmacología de una determinada vacuna en varones sanos universitarios, quizá los resultados no fueran aplicables a personas de otro sexo o edad.

Por otra parte, los criterios de exclusión son útiles para hacer el estudio más homogéneo, pero limitan la aplicabilidad de los resultados. Si se excluyen las formas leves de la enfermedad, el estudio efectuado con formas graves no será aplicable a casos leves-moderados. Así, por ejemplo, un estudio que incluyó solo pacientes hospitalizados no tendrá conclusiones válidas para el manejo de pacientes ambulatorios.

Como destaca Rebagliato[90], es importante tener presente entonces que los criterios de selección del estudio normalmente reflejan alguno de los siguientes aspectos **(tabla 15)**.

Tabla 15. Aspectos que se reflejan según el criterio de selección del estudio	
Costos	Algunos criterios reflejan restricciones de coste. Por ejemplo, si excluimos a personas que no hablan español no significa que no nos interese ese colectivo, sino que no tenemos posibilidades de contratar intérpretes
Problemas de orden práctico	Problemas para incluir a personas de áreas rurales, con dificultad de seguimiento (transporte)
Posibilidad de participar en un estudio	Puede ser necesario excluir de un estudio a pacientes con cáncer en fase terminal
Diseño del estudio	A veces es más adecuado definir una muestra más homogénea para controlar las variables extrínsecas (nivel cultural, socioeconómico, educativo)

¿Cómo se hizo el muestreo?

Resulta lógico pensar que para los investigadores de cualquier trabajo va a ser imposible estudiar a toda la población elegible, ya sea por una cuestión de tiempo, dinero u otros recursos. Al no estudiar la población en su conjunto los investigadores han de elegir una muestra. El término muestreo se refiere al proceso de selección de una parte o subconjunto de la población para que represente al conjunto[91].

Una muestra es, por tanto, una parte o subconjunto de la población, que mediante técnicas estadísticas nos permite, a partir de los datos muestrales, inferir los resultados al resto de la población. Las unidades que constituyen las muestras y las poblaciones se conocen como elementos o unidades de análisis. El elemento es, pues, la unidad básica acerca de la cual se recaba información.

El objetivo principal a la hora de evaluar una muestra es la representatividad. Es decir, mediante el muestreo los autores han de haber elegido una muestra que sea representativa, cuyas características clave se correspondan estrechamente con las de la población de la que parte y que, por tanto, tenga el mismo grado de diversidad que la población. Por desgracia, el concepto de representatividad tiene mucho de intuición,

y no sabemos si la muestra es representativa, puesto que la representatividad está vinculada con el grado de variabilidad de la población. Lo que sí que sabemos es que los diseños de muestreo probabilísticos, en los que interviene el azar, normalmente producen muestras representativas, garantizando que cada unidad de población tenga una probabilidad específica de ser seleccionada. Por el contrario, cuando las unidades de población se eligen utilizando métodos en los que no interviene el azar *(muestreo no probabilístico),* no es posible estimar la probabilidad de que cada elemento haya sido incluido en la muestra y no todos los elementos tienen la posibilidad de ser incluidos.

Por su mayor interés y aplicación en medicina, a continuación se resumen las características principales de los diseños de muestreo probabilísticos **(tabla 16)**.

¿Es adecuado el tamaño de la muestra?

Tras haber abordado los distintos tipos de diseños, algo que preocupa a los autores de un artículo científico es saber cuántos individuos debe tener la muestra, es decir, si el tamaño de la muestra elegida fue el adecuado: representativa de la población objeto de estudio. Por ello, al leer un artículo científico no debemos olvidar las dificultades que han de vencer los investigadores, que han de calcular el tamaño de la muestra para conocer algo que a priori se desconoce, pero para lo que, sin embargo, deben partir de un conocimiento de su valor,

Tabla 16. Tipos de diseño de muestreo probabilístico (modificada de Robledo Martín[92,93])

Tipo de muestreo	Ventajas	Inconvenientes
Muestreo aleatorio simple		
La probabilidad es la misma para todos los elementos	El más sencillo conceptual y estadísticamente El que produce muestras más representativas, puesto que solo interviene el azar y, por tanto, no está la muestra sujeta a sesgos del investigador	Requiere una lista enumerada de todas las unidades de población La posible dispersión de los sujetos seleccionados, por lo que contactar con ellos será costoso en tiempo y dinero Los subgrupos más minoritarios pueden no estar representados en la muestra, si es pequeña
Muestreo sistemático		
Selección de individuos según una regla o proceso periódico; se calcula la constante K de muestreo, dividiendo el total de la población elegible entre el tamaño de la muestra deseado K = N/n. Tras elegir la primera unidad al azar, se va sumando consecutivamente *K* hasta completar la *n*	No necesita tener la lista de la población cerrada de antemano para seleccionar la muestra	No es aconsejable en las situaciones en las que las unidades de muestreo están ordenadas por algún criterio periódico y la K puede coincidir con ello

Tabla 16. Tipos de diseño de muestreo probabilístico (modificada de Robledo Martín[92,93]) *(continuación)*

Tipo de muestreo	Ventajas	Inconvenientes
Muestreo estratificado		
Se divide la población en subgrupos o estratos, de acuerdo con una variedad de atributos (edad, sexo, nivel socioeconómico, ocupación, etc.)	Permite estimar con determinada precisión cómo se comporta la variable de la investigación en subgrupos pequeños dentro de la población	Requiere mayor información que en el muestreo aleatorio simple, ya que se necesita una lista enumerada de la población existente en cada uno de los estratos
Muestreo conglomerado		
La población la dividimos en agrupaciones naturales de individuos como hospitales, colegios, provincias, etc. a los que denominamos conglomerados	Idóneo cuando la población de referencia está muy dispersa o cuando no se dispone de un listado de individuos, pero sí de las agrupaciones o conglomerados	Requiere muestras de mayor tamaño Imposible conocer de antemano el tamaño de la muestra que se va a obtener (si el tamaño entre conglomerados es muy diferente, la variabilidad será elevada)
Todas las unidades de población tienen una probabilidad de ser elegidas. Esa probabilidad es conocida de antemano		

que en la mayoría de los casos lo obtendrán a partir de otros estudios. Deben definir el efecto clínicamente importante que quieren ser capaces de detectar y calcular entonces qué

tamaño de muestra confiere al estudio la potencia estadística suficiente para demostrarlo.

Querido lector, seguro que en este punto estaría encantado de que le proporcionara una fórmula que le dé un "número mágico" de sujetos que formen la muestra de pacientes que debe ser incluida en un estudio epidemiológico. Le adelanto que para comprender su desarrollo necesitamos conocer algunos conceptos estadísticos que desarrollaremos a continuación. Por el momento, desechemos ese concepto de "número mágico", y tengamos en cuenta algunas consideraciones.

El tamaño de la muestra de pacientes que se van a incluir en un estudio depende de cuatro factores: la homogeneidad de las poblaciones participantes, la cuantía de las diferencias que se pretende poner de manifiesto y los errores tipo I y II que se consideran aceptables[94]. Las muestras de pacientes más homogéneos confieren mayor probabilidad de detectar pequeñas diferencias. Si la muestra de participantes es más heterogénea, su tamaño ha de ser mayor para poder observar diferencias entre ambos grupos. Análogamente, cuanto mayores sean las diferencias que se desee demostrar entre dos intervenciones, menor será el número de pacientes que

se necesita. No nos debería extrañar que sea necesaria una muestra más amplia para descartar que en una enfermedad grave el tratamiento disminuye la mortalidad del 20% al 18%, que para descartar que lo reduce del 20% al 5%.

Hay que tener en cuenta también que el tamaño de la muestra está en función del error máximo de muestreo que se esté dispuesto a admitir al estimar una variable. La decisión de determinar a priori el error máximo permitido no deja de ser esencialmente subjetiva. El error de tipo I o falso positivo corresponde, según la definición usual a la probabilidad de rechazar la hipótesis nula (Ho) cuando esta es verdadera[95]; en otras palabras, en afirmar que existen diferencias entre los dos grupos de estudio, cuando en realidad no las hay. A la probabilidad de cometer un error de tipo I se le denomina "α", y equivale a la probabilidad de encontrar un resultado falso positivo, porque el investigador llega a la conclusión de que existe una diferencia entre las hipótesis cuando en realidad no existe (p. ej., se considera que el paciente es portador del VIH, a pesar de que está sano; H_o: el paciente no está infectado por el VIH).

El *error de tipo II* también llamado error de tipo "β" (β es la probabilidad de que exista este error) o *falso negativo*, se comete cuando el investigador no rechaza la hipótesis nula siendo esta falsa en la población. Es equivalente a la probabilidad de un resultado falso negativo, ya que el investigador llega a la conclusión de que ha sido incapaz de encontrar una

diferencia que existe en la realidad. El poder o potencia del estudio representa la probabilidad de observar en la muestra una determinada diferencia o efecto, si existe en la población. Es el complementario del error de tipo II (1-β). Siguiendo con el mismo ejemplo, un error de tipo II implicaría que el paciente no tiene VIH cuando no es así; desde luego, un diagnóstico peligroso.

Cuanto más complejo es el diseño de muestreo que se use, mayor será la muestra que se necesita, puesto que el efecto diseño hace que aumente el tamaño de la muestra para conseguir el mismo grado de precisión. Un tamaño de muestra pequeño puede dar lugar a un error de tipo II, mientras que el error de tipo I es más bien debido al azar.

Finalmente, el tamaño de la muestra condiciona de un modo importante la posibilidad de detectar el efecto de la intervención. Si la muestra es pequeña puede que no se detecte el efecto existente, mientras que si la muestra es grande puede conseguirse significación estadística, con diferencias de efecto que son poco importantes clínicamente.

En resumidas cuentas, concluiremos esta sección con una frase del epidemiólogo Keneth Rothman: "... El problema de determinar el tamaño de la muestra más adecuado no es de naturaleza técnica... sino que ha de encararse mediante el juicio, la experiencia y la intuición".

2.2.8. ¿Son adecuadas las características metodológicas?

Para responder a esta pregunta no se requiere ser un experto en estadística, pero sí se requiere tener algunos conceptos claros sobre cómo y cuándo aplicar un determinado tipo de test estadístico. Existen elementos estadísticos a los que se debe poner especial atención, que deben estar explícitamente detallados en la sección de métodos del artículo: *a)* el cálculo del tamaño muestral; *b)* el análisis según intención de tratar, y *c)* la duración estimada del seguimiento y el número estimado de pacientes que lo completarán.

El cálculo del tamaño de la muestra de pacientes de un estudio

Una vez valorados los objetivos y el tipo de estudio, el lector debe concentrarse en el tamaño de la muestra seleccionada de individuos. Además, ha de preguntarse si el número de pa-

cientes incluidos en el estudio es suficiente para demostrar con una probabilidad razonable la existencia de una diferencia estadísticamente significativa entre las muestras del estudio y si dicha diferencia existe realmente en la población de la cual se han extraído las muestras.

La mayor parte de los estudios clínicos controlados tienen como objetivo determinar si el nuevo tratamiento produce un resultado mejor que el placebo o el tratamiento estándar. Para decidir si el ensayo es viable, es preciso estimar el tamaño de la muestra necesaria. En otras palabras, los investigadores deben averiguar cuántos pacientes es necesario estudiar para tener una probabilidad razonable de demostrar una diferencia estadísticamente significativa entre el nuevo tratamiento y el placebo o el tratamiento estándar.

Cuando nos preguntamos por la idoneidad del tamaño de la muestra, es preciso distinguir entre los estudios de casos y controles, por un lado, y los ensayos clínicos aleatorios y los estudios de cohortes por el otro. Mientras que en los estudios de casos y controles el desenlace es una característica del paciente, en los estudios de cohortes y en los ensayos clínicos aleatorios el desenlace es una enfermedad. Por consiguiente,

en un estudio de casos y controles, estaríamos interesados en conocer la magnitud de la verdadera diferencia entre ambos, que sin tanta limitación por el tamaño de la muestra, es probable demostrar como estadísticamente significativa. Al diseñar un estudio de cohortes o un ensayo clínico aleatorizado, uno está interesado en saber cuán pequeña debe ser una verdadera diferencia en la probabilidad de desarrollar una enfermedad, para que sea probable demostrar que es estadísticamente significativa, dado el tamaño de la muestra empleado en el estudio.

Como vimos en el apartado anterior, las respuestas a estas cuestiones dependen de la magnitud de los errores de tipo I y II que tanto el lector como el investigador estén dispuestos a tolerar. El error de tipo I es la probabilidad de demostrar una diferencia estadísticamente significativa en las muestras cuando no existe una verdadera diferencia entre los tratamientos en las poblaciones. El nivel α correspondiente al tipo I se sitúa normalmente en el 5%. El error de tipo II, sin embargo, es un tema por discutir. Recordemos que el error de tipo II es la probabilidad de aceptar una hipótesis nula que es falsa, es decir, de no ser capaces de demostrar una diferencia estadísticamente significativa, aun existiendo dicha diferencia entre el grupo de estudio y el de control. El objetivo deseado por la mayoría de los investigadores para el error de tipo II es que no sea superior al 20%. Un error del 20% también se denomina potencia estadística del 80%. La potencia del 80% implica

que existe una probabilidad del 80% de demostrar una diferencia estadísticamente significativa, cuando existe realmente. Existen numerosos estudios publicados con baja potencia, probablemente porque para los autores resultó mucho más difícil de lo estimado reclutar a los pacientes. Esto lleva al error de tipo II o β, que es la conclusión errónea de que la intervención no tuvo efecto.

Otros factores que también influyen en la decisión del tamaño de la muestra de pacientes son, por un lado, el porcentaje de individuos en el grupo de control que se espera que experimentarán el desenlace adverso estudiado –muerte o desarrollo de la enfermedad– y, por otro, la magnitud de la diferencia perseguida por el investigador, o lo que es lo mismo, la mejora en el desenlace entre los miembros del grupo de estudio que se pretende demostrar como estadísticamente significativa. A pesar del deseo de demostrar una diferencia estadísticamente significativa, incluso para cambios reales pequeños, es necesario que los investigadores decidan la magnitud de la diferencia que sería considerada clínicamente importante. Cuanto menor sea la diferencia que se pretende observar entre el grupo de control y el de estudio, mayor será el tamaño de la muestra requerido.

En los ensayos clínicos controlados generalmente se utilizan muestras de 100. Esta es una estimación del número de individuos necesarios en cada grupo cuando la probabilidad de un efecto adverso es sustancial y los investigadores esperan

reducirla a la mitad con el nuevo tratamiento, mientras mantienen la magnitud del error de tipo II por debajo del 20%.

Si aceptamos un error de tipo I del 5%, y del 10% para el error de tipo II, y utilizamos tablas estadísticas estándares, se pueden extraer las siguientes conclusiones sobre el tamaño de la muestra **(tabla 17)**.

Tabla 17. Conclusiones sobre el tamaño de la muestra

Tamaño de la muestra (grupo de control/grupo de estudio)	100	250	500
Frecuencia del desenlace en el grupo de control (p. ej., mide el riesgo de sepsis neonatal con el tratamiento estándar)	≥ 20%	≥ 20%	≥ 10%
Frecuencia del desenlace en el grupo de estudio (p. ej., efecto de un nuevo fármaco sobre el riesgo de sepsis neonatal)	≤ 5%	≤ 10%	≤ 5%

Siguiendo este ejemplo, un tamaño de muestra de 500 individuos en cada grupo es una cifra grande para un ensayo clínico controlado. Esta es la cifra aproximada que necesitamos,

si queremos que nuestra investigación tenga una potencia estadística para demostrar una diferencia estadísticamente significativa cuando la verdadera diferencia entre las frecuencias de desenlace en la población es solo del 10% frente al 5%.

Estas estimaciones son útiles para el lector de la literatura médica, porque le permiten estimar si el estudio tiene una posibilidad real de demostrar una significación estadística a partir del tamaño de su muestra. Ahora bien, cuando use estas orientaciones, recuerde que, incluso con una potencia estadística muy elevada (1-$\beta \geq 90\%$), una muestra concreta extraída de poblaciones en las que existen verdaderas diferencias aún puede ser insuficiente para detectar una diferencia estadísticamente significativa.

La importancia de la intención de tratar y del seguimiento

Idealmente, todos los pacientes incluidos en la muestra de un ensayo clínico debieran recibir la intervención a la que fueron asignados y ser seguidos hasta el final del estudio. Por desgracia, todos sabemos que esto no es así, y frecuentemente nos encontramos frente a dos situaciones que pueden introducir un sesgo: *a)* no todos los pacientes reciben la intervención asignada, ya sea tratamiento activo o placebo, y *b)* no

siempre el seguimiento de los pacientes es completo, es decir, se desconoce si ocurrió el *outcome* de interés en algunos de los pacientes incorporados al estudio.

Sin duda, estas dos situaciones comprometen la validez interna y externa de un ensayo. Revisemos cuáles son las estrategias que nos permitirán evaluar estos criterios de validez y cómo el investigador podría disminuir el posible sesgo introducido por ellas.

El principio del análisis según la intención de tratar

La verificación del correcto análisis de todos los pacientes participantes en el estudio incluye la comprobación de un seguimiento completo y un análisis de los resultados por intención de tratar. El principio de *intención de tratar* es una forma de analizar los resultados que considera a todos los individuos pertenecientes al estudio, conforme al grupo al cual fueron asignados originalmente, aunque no hayan cumplido con el protocolo[96]. Esto permite mantener un equilibrio entre los factores pronósticos conocidos y desconocidos, disminuyendo la probabilidad de sesgar los resultados.

En los ensayos clínicos, como en la práctica clínica diaria, se producen con cierta frecuencia desviaciones del protocolo.

Esto ocurre cuando un paciente que fue asignado aleatoriamente a recibir un tratamiento deja de tomarlo, lo sigue parcialmente o recibe en realidad, por distintos motivos, el tratamiento de otro grupo.

Ante estas circunstancias cabrían teóricamente tres posibilidades: *a) excluir* al paciente del análisis final; *b)* analizarlo en el grupo del tratamiento que realmente realizó, aunque fuera asignado a otro (análisis por tratamiento), o *c)* analizarlo en el grupo de tratamiento al que fue asignado aleatoriamente, aunque realmente no lo recibiera *(análisis por intención de tratar)*. Paradójicamente, esta última opción es el enfoque más correcto para reducir la probabilidad de sesgo y así mejorar la validez interna del ensayo clínico aleatorizado. Además, cabe pensar que si la intervención demuestra ser efectiva con un análisis por intención de tratar en pacientes que no recibieron la intervención del estudio, entonces podemos estar seguros de que tal intervención es aún más efectiva en los pacientes que sí se adhirieron a ella. Las principales ventajas y limitaciones del análisis por intención de tratar se resumen como sigue[97] **(tabla 18)**.

Los artículos científicos sobre un estudio en investigación clínica suelen incluir la coletilla "El análisis se realizó por intención de tratar" en el apartado material y métodos, en la sección de métodos estadísticos.

Tabla 18. Principales ventajas y limitaciones del análisis por intención de tratar

Ventajas	Limitaciones
Se conservan las propiedades conseguidas con la asignación aleatoria: la homogeneidad de los grupos. Y mantiene el equilibrio de los factores pronóstico relevantes para el resultado	Menor magnitud del efecto de la intervención, debido al efecto de los sujetos que no se adhieren al protocolo
Visión más real y menos sesgada de la magnitud del efecto de una determinada intervención	No válido para los ensayos de equivalencia (no hay diferencia entre las intervenciones), porque puede favorecer la equivalencia
Se aproxima a la realidad de la práctica clínica (efectividad), en la que los cambios o incumplimiento de los tratamientos son habituales	No es el más adecuado en estudios que evalúan los efectos adversos de una intervención

¿Fue completo el seguimiento?

Otro elemento de sesgo íntimamente relacionado con el principio de intención de tratar, pero que este no puede vencer, es

el seguimiento incompleto de pacientes. Los estudios tienen una tasa variable de pérdida de seguimiento de pacientes que será mayor en aquellos con mayor tiempo de seguimiento. Durante el desarrollo de un ensayo clínico aleatorizado, las pérdidas posaleatorización incluyen:

- **Los abandonos** *(drop out)* son los pacientes incluidos en un ensayo clínico que no quieren o no pueden continuar en el estudio ni en el seguimiento.

- **Las retiradas** *(withdrawal)* son los pacientes obligados a abandonar el estudio por presentar determinadas circunstancias especificadas en el protocolo (como la aparición de nuevos criterios de exclusión, reacciones adversas graves, la aparición de una enfermedad concomitante o por la sensación de falta de mejoría), pero en los que sí es factible su seguimiento.

- **Las pérdidas de seguimiento** *(loss to follow-up)* son los pacientes incluidos en el estudio, que no realizan los controles indicados y de los que no es posible obtener datos de su seguimiento. Por ejemplo, porque el paciente se cambia de domicilio y deja de pertenecer al hospital o centro ambulatorio donde se desarrolla el estudio.

- **Las desviaciones del protocolo** son los pacientes que por distintas circunstancias no siguen el protocolo asignado (incumplimientos terapéuticos, cambio de brazo por error, etc.).

¿Cuál es entonces la magnitud de pérdida posaleatorización aceptable para considerar que un ensayo clínico es válido? Algunos autores aceptan el 20%, otros menos del 10%. Esta variabilidad de opciones se debe a que el porcentaje de pérdida aceptable varía con relación a la tasa de *outcome* de interés que ocurrió en el estudio[98]. Así, en un estudio con un 2% de mortalidad en el grupo de estudio y un 4% de mortalidad en el grupo de control, una pérdida tan solo del 2% de los individuos aleatorizados puede alterar significativamente los resultados.

Comprobadas las preguntas esenciales que debe cumplir todo artículo científico, deben verificarse otros detalles que aportan rigor y calidad metodológica a los estudios: el análisis estadístico de los resultados. Y esto es lo que te proponemos leer a continuación.

2.3. El cumplimiento de la cuestión estadística

La estadística aplicada a la investigación médica persigue tres finalidades: *a)* sintetizar numerosas mediciones en un número limitado de datos manejables; *b)* realizar estimaciones e inferencias a partir de las muestras extraídas de poblaciones, teniendo en cuenta la influencia del azar, y *c)* ajustar los datos según la influencia de las variables de confusión en esas estimaciones e inferencias[99].

Cronológicamente, para utilizar la estadística en la investigación médica, es preciso elegir un método estadístico apropiado, las mediciones de la investigación deben ser manipuladas de acuerdo con el método seleccionado y, por último, los resultados de estas manipulaciones han de interpretarse correctamente.

Antes de continuar con la lectura, permítame que comparta la siguiente reflexión:

No se pretende que la información facilitada a continuación pueda reemplazar la participación de un profesional estadístico en las fases de planificación, ejecución e interpretación de cualquier proyecto de investigación médica. Tampoco se discutirá a fondo en este libro sobre las manipulaciones de los datos que son necesarias para producir los resultados estadísticos. No es preciso tener ese nivel de conocimientos para poder evaluar por qué se selecciona un método determinado y cómo podemos interpretar los resultados de su aplicación. Lo que se persigue con esta parte del libro es proporcionar las herramientas necesarias para que los lectores de la literatura de investigación sepan valorar la sección de "métodos estadísticos", de tal forma que el análisis, la interpretación y la extrapolación de los resultados de la investigación se comprendan en su totalidad.

2.3.1. Objetivos de la cuestión estadística

Síntesis de las medidas

Como se acaba de afirmar, una de los objetivos de la estadística consiste en resumir grandes cantidades de datos a un número reducido y manejable. Para cumplir con esto, hemos de tener en cuenta que las medidas realizadas en los sujetos del estudio son una muestra de un grupo más numeroso de individuos al que todos denominamos *población*. Una *población estadística* se define como el conjunto de todas las mediciones posibles de las cuales se selecciona una muestra.

Si representamos gráficamente la frecuencia con que aparecen los distintos valores de una variable en la población, obtendremos una distribución poblacional **(figura 10)**. Los métodos estadísticos ofrecen una medida sintética de la distribución poblacional, en lugar de su descripción gráfica. Cada tipo de distribución poblacional tiene un número limitado de valores sintéticos, denominados *parámetros*, que se utilizan para describir completamente la distribución concreta de las mediciones. Por ejemplo, para describir íntegramente una *distribución gaussiana*, se necesitan dos parámetros: la *media* (la posición de la distribución en una escala continua o, más concretamente, su *centro de gravedad*) y la *desviación estándar* (la dispersión de la distribución, dado que indica cuán alejados de la media se encuentran los valores individuales).

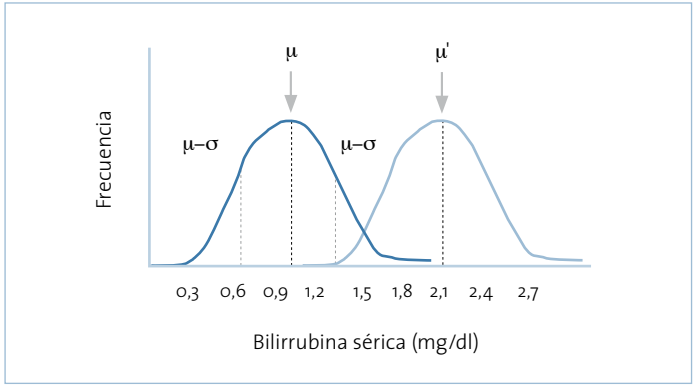


Figura 10. Distribución gaussiana hipotética de la concentración de bilirrubina sérica con una media de 0,9 mg/dl (o de 1,5) y una desviación estándar de 0,3 mg/dl. Las líneas discontinuas indican los valores iguales a la media ± la desviación estándar.

La figura 10 muestra una distribución gaussiana de la concentración de bilirrubina sérica con la media (μ = 0,9 mg/dl), como medida de posición de la distribución, y la desviación estándar (σ = 0,3), como medida de dispersión. Para demostrar lo que queremos decir con la posición de una distribución, supongamos que la media de la concentración sérica de bilirrubina en la población es de 1,5 ml/dl, en lugar de 0,9 mg/dl.

Estimación e inferencia

La *estadística descriptiva* y la *teoría de la probabilidad* van a ser los pilares de un nuevo procedimiento: *la estadística infe-*

rencial, con la que se va a estudiar el comportamiento global de un fenómeno. La probabilidad y los modelos de distribución junto con las técnicas descriptivas constituyen la base de una nueva forma de interpretar la información suministrada por una parcela de la realidad que interesa investigar. Los métodos básicos de la estadística inferencial son la *estimación* y el *contraste de hipótesis*, que desempeñan un papel fundamental en la investigación[100].

En la mayoría de las investigaciones resulta imposible estudiar a todos y cada uno de los individuos de la población, ya sea por el coste que supondría o por la imposibilidad de acceder a ello. Mediante la técnica inferencial obtendremos conclusiones para una población no observada en su totalidad, a partir de estimaciones o resúmenes numéricos extraídos de una muestra de dicha población. Por tanto, el esquema que se debe seguir es el que aparece en la **figura 11**.

En definitiva, la idea es, a partir de una población se extrae una muestra (véase apartado 2.2.8.), con la que se generan datos numéricos que se van a utilizar para generar *estadísticos* con los que realizar *estimaciones* o *contrastes poblacionales*. Estas estimaciones concretas son los estadísticos que usamos para resumir grandes cantidades de mediciones en unas pocas y más manejables.

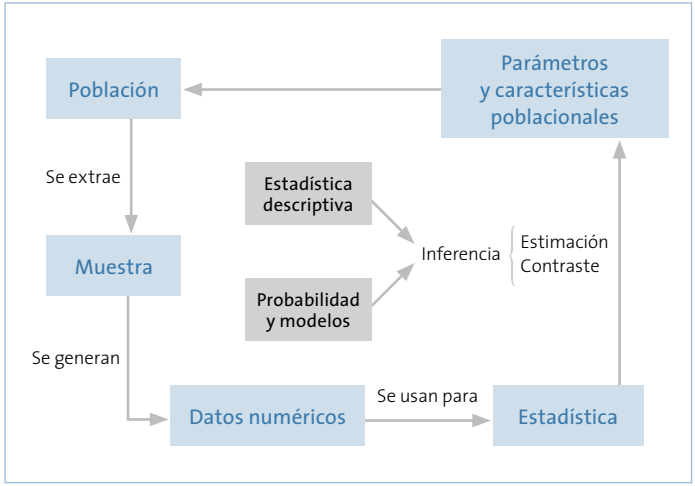


Figura 11. Esquema de la estadística diferencial aplicada a un fenómeno.

Lo que podemos saber, por tanto, es cuánto se espera que varíe el estadístico en relación con el valor hipotético del parámetro poblacional sobre la base de la variabilidad del azar entre las muestras aleatorias. Este conocimiento constituye la base de la inferencia estadística o de las pruebas de significación estadística, concepto al que volveremos más adelante.

Existen dos formas de estimar parámetros: la *estimación puntual* y la *estimación por intervalo de confianza*.

- Mediante la **estimación puntual**, se busca un valor concreto para el parámetro, a partir de los datos muestrales. En pocas palabras, se trata de una fórmula que depende de los valores obtenidos de una muestra para realizar estimaciones. Lo que se pretende obtener es el valor exacto de un parámetro. Por ejemplo, si se quiere estimar la glucemia media de un determinado grupo de individuos, puede extraerse una muestra y ofrecer como estimación puntual la concentración media de glucosa de los individuos de la muestra[100].

- Mediante la **estimación por intervalo** o **intervalo de confianza**, se determina un intervalo dentro del cual se encuentra el valor del parámetro, con una probabilidad determinada. Se trata de unos límites o margen de variabilidad que damos al valor estimado, para poder afirmar, bajo un criterio de probabilidad, que el verdadero valor no los rebasará ni por encima ni por debajo del rango. Si bien es posible calcular el intervalo de confianza para cualquier porcentaje de confianza entre 0 y 100, el del 95% es el utilizado con más frecuencia. Un intervalo de confianza del 95% nos permite tener una *confianza* del 95% de que el valor de la población (parámetro) se halla dentro de ese intervalo. A partir este valor, también conocido como nivel de confianza (expresado como $1-\alpha$), podemos calcular

el nivel de significación estadística de la prueba (expresado como α y calculado en forma de tanto por uno). Es decir, un nivel de confianza del 95% se corresponde con un nivel de significación de 0,05. El nivel de significación es la probabilidad de fallar en nuestra estimación, esto es, indica la diferencia entre la certeza (1) y el nivel de confianza (1-α).

El *contraste de hipótesis* es una prueba de significación estadística que indica el proceso mediante el cual decidimos si una proposición respecto de la población, debe ser aceptada o no[100]. Esta proposición es lo que se denomina *hipótesis estadística*. Es una regla de decisión que nos dice cuándo aceptar y rechazar las hipótesis o, dicho de otro modo, con esto vemos si los datos de una muestra son compatibles con los de la población. Una hipótesis estadística, por tanto, es una afirmación acerca de la distribución de la población.

El contraste de hipótesis estadístico se basará en la información proporcionada por la muestra. Se estudian dos hipótesis: H_o (hipótesis nula), H_1 (hipótesis alternativa), de manera que el investigador divide los resultados de las muestras en dos zonas; una zona de rechazo y otra de aceptación, de modo que según como obtengamos el resultado, aceptaremos o rechazaremos la hipótesis nula. Si hay coincidencia, dentro de un margen de error admisible entre la hipótesis planteada (hipótesis nula, H_o) y la información de la muestra, entonces la mantendremos como cierta; en caso contrario, la rechaza-

remos. Si rechazamos la hipótesis, queremos indicar que los datos de la muestra ofrecen cierta evidencia sobre su falsedad. Rechazar una hipótesis es sustituirla por otra que sea capaz de explicar la realidad observada en la muestra (hipótesis alternativa, H_1).

Ahora bien, nunca podemos afirmar que una hipótesis es verdadera o falsa, ya que para ello deberíamos tener observaciones de toda la población. Por tanto, al realizar un contraste y tomar una decisión siempre cabe la posibilidad de equivocarnos. Existen dos tipos de errores asociados a cualquier contraste: error tipo I, que tiene lugar cuando rechazamos H_0 siendo cierta, y error tipo II, que ocurre si aceptamos H_0 siendo falsa. La descripción de ambos errores ya se ha tratado anteriormente en esta obra (v. tabla 17, sobre la importancia del tamaño de la muestra).

Como lector de la literatura clínica, usted encontrará cada vez más valores observados e intervalos de confianza en la sección de resultados de los artículos científicos. Esto es una gran ayuda, porque permite hacerse una idea sobre los datos y le permite sacar, además, sus propias conclusiones sobre la importancia clínica de la dimensión o de la fuerza de la estimación puntual observada. Por último, para aquellos que desean convertirlo al formato tradicional de las pruebas de significación estadística para contraste de hipótesis, muchas veces se puede realizar un cálculo aproximado para determinar si los resultados son estadísticamente significativos con un valor p de 0,05 o menor.

Ajuste de los datos

El investigador está obligado a comparar las características de los individuos del grupo de estudio con las del grupo de control, para determinar si difieren en alguna de ellas. Si hay diferencias entre los grupos, aunque no sean estadísticamente significativas, el investigador debe considerar si aquellas han influido en los resultados. Como ya se definió con anterioridad en el apartado correspondiente a la minimización de los sesgos (v. apartado 2.2.6.), las características que indican

diferencias entre los grupos y que pueden influir en los resultados del estudio son variables de confusión potenciales. Estas variables pueden ser el resultado de un sesgo de selección en los estudios de casos y controles o de cohortes, o de diferencias aleatorias en los tres tipos de estudios básicos –estudios de cohortes, de casos y controles y ensayos clínicos aleatorizados–. Si el investigador detecta una variable de confusión potencial, está obligado a tenerla en consideración en el análisis mediante un proceso denominado *ajuste de los datos*.

Mediante el ajuste de los datos, el investigador separa en grupos a los sujetos que poseen niveles diferentes de la variable de confusión, compara después los grupos con el mismo nivel de la variable de confusión y, finalmente, analiza si existe una asociación entre la exposición y la enfermedad. Así, por ejemplo, si la edad es una variable de confusión potencial, los grupos de estudio y de control habrán de ser agrupados por edades en diversas categorías y determinar así si hay diferencias al comparar los grupos de edades similares.

Para realizar un ajuste de los datos se dispone de técnicas estadísticas conocidas como *métodos multivariantes* para ajustar los datos según una o más variables al mismo tiempo[99]. En el análisis multivariante tenemos una variable dependiente y dos o más independientes. Estas variables independientes se pueden medir en la misma escala o en diferentes. El uso de los métodos multivariantes para analizar los datos

de la investigación médica presenta ciertas ventajas, pero la más importante quizá sea que permite investigar la relación entre una variable dependiente y una independiente, mientras se controla o se ajusta según el efecto de otras variables independientes. Por ese motivo, los métodos multivariantes se utilizan para cumplir con la tercera finalidad de la estadística en el análisis de los resultados de la investigación médica: **ajustar según la influencia de las variables de confusión**. Por ejemplo, si queremos estudiar la tensión arterial diastólica en sujetos que reciben distintas dosis de un fármaco antihipertensivo, podríamos desear controlar la influencia de la edad y del sexo como posibles variables de confusión. Para ello, utilizaríamos un análisis multivariante con la tensión arterial diastólica como variable dependiente y la dosis, la edad y el sexo como variables independientes.

En resumen, mediante la síntesis de datos y las técnicas de asignación al azar y de ajuste, los investigadores pueden crear grupos de estudio y de control que sean comparables excepto en los efectos del tratamiento administrado. Por eso, cuando veamos en un artículo en el que los investigadores concluyen que existen diferencias notables y estadísticamente significativas en el desenlace, debemos estar seguros de que estas diferencias están asociadas al tratamiento.

Dicho de otro modo, como decía uno de mis dos mentores durante mi formación de posgrado en la Universidad:

"... pues sí, queridos alumnos, vosotros, como futuros investigadores, habréis de saber que hay que ser muy cauto con lo que nos dice la estadística,... porque la estadística es como un biquini, es mucho lo que enseña, pero es esencial lo que esconde..."

2.3.2. Pruebas de significación estadística

Desarrollar cualquier estudio clínico tiene como objetivo poner de manifiesto la existencia de asociación entre variables. Para ello, como acabamos de ver, se usa la inferencia estadística, proceso que le permite al investigador establecer conclusiones sobre una población en función de los resultados obtenidos en una muestra que se considera representativa. Sin embargo, cualquier investigador ha de enfrentarse a menudo con la duda de si se hubieran obtenido unos resultados similares con toda la población, o si el azar pudo haber producido unos resultados de excepción en la muestra elegida. Lamentablemente, no existe un método directo para responder a esta pregunta. En su lugar, los investigadores están obligados a poner a prueba sus hipótesis de estudio, empleando un método indirecto de prueba por eliminación. Este método se conoce como *prueba de significación estadística*[101].

El concepto *significación estadística* se relaciona con la necesidad de *probar hipótesis*. Este proceso se realiza utilizando *pruebas de hipótesis*, las que permiten cuantificar hasta qué punto la variabilidad de la muestra en estudio es responsable de los resultados obtenidos en el estudio.

La hipótesis nula plantea la ausencia de diferencias significativas (no hay asociación entre variables), mientras que la hipótesis alternativa afirma la presencia de diferencias (sí hay asociación entre las variables). Las pruebas que permiten analizar estas diferencias son conocidas como pruebas de significación estadística y cuando son significativas ($p < 0,05$) conllevan el rechazo de la hipótesis nula y la aceptación de la hipótesis alterna.

Entonces, la estadística nos permite decidir sobre que hipótesis debemos elegir, lo que será con el nivel de seguridad que previamente se haya establecido por el equipo de investigación, que habitualmente en clínica es el 95%. El *valor de p* que indica que la asociación es estadísticamente significativa ha sido arbitrariamente aceptado por consenso. Y en clínica se admite 0,05. Dicho en otros términos, tenemos una seguridad del 95% de que la asociación que estamos estudiando no se deba al azar. Si quisiéramos trabajar con un margen de seguridad aún mayor –del 99%–, lleva implícito una $p < 0,01$.

Las pruebas de significación estadística han de seguir un esquema de desarrollo **(figura 12)**, que siempre empieza con la formulación de una hipótesis de estudio que afirma que existe una asociación entre factores en la población elegible. Al realizar una prueba de significación, inicialmente se supone que

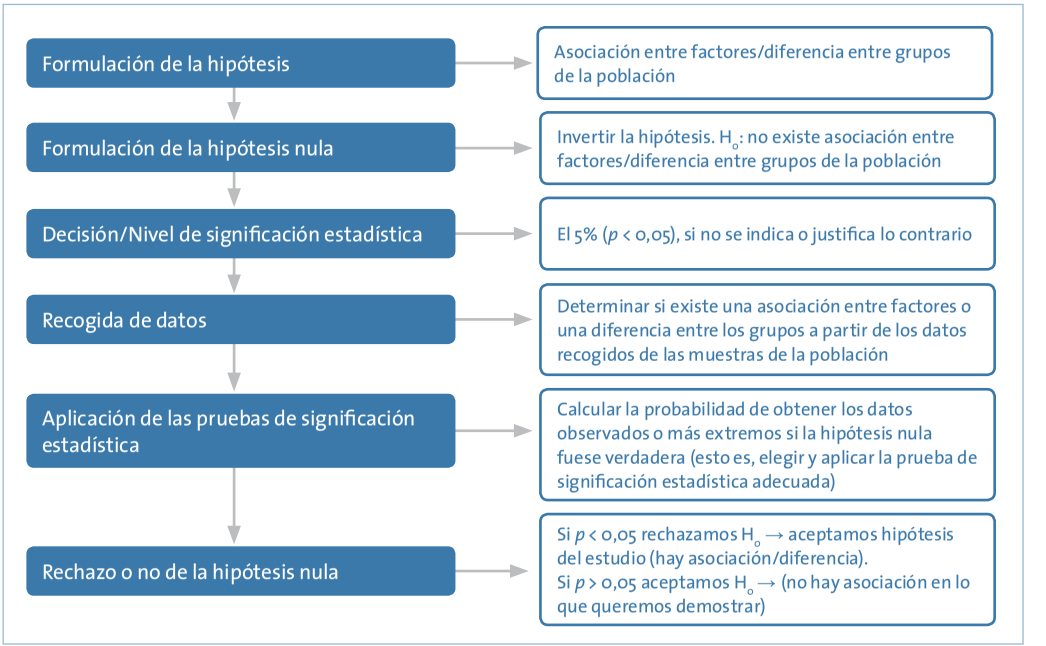


Figura 12. Esquema del diseño de una prueba de significación estadística.

la hipótesis de estudio es falsa, y se formula entonces una hipótesis nula según la cual no existe dicha asociación o diferencia en la población. Los métodos estadísticos se emplean entonces para calcular la probabilidad de obtener los resultados observados en la muestra estudiada. Cuando solo existe una pequeña probabilidad de obtener los resultados observados si la hipótesis nula fuese verdadera, los investigadores pueden entonces rechazar la aseveración de que la hipótesis nula es cierta y, con ello, la hipótesis nula. Al hacerlo, aceptan por eliminación la existencia de su única alternativa, la de una asociación o diferencia entre las variables estudiadas en la población.

Las pruebas de significación estadística son métodos de contraste de hipótesis que sirven para valorar los efectos del azar en los resultados de una investigación, ayudándonos a medir la *fuerza de una asociación* observada[101]. En el ensayo clínico y en el estudio de cohorte el *riesgo relativo* es la medida básica de la fuerza de una asociación. Por otro lado, en el estudio de casos y controles es la *razón de momios* (*odds ratio* de la literatura sajona) la que informa de la fuerza de la asociación.

Por asociación entendemos que un factor, con frecuencia llamado *factor de riesgo*, nos permite cuantificar si existe un aumento o una disminución del riesgo de padecer una enfermedad. En ciencias de la salud, el término *riesgo* constituye una medida de probabilidad estadística de que los individuos expuestos a ciertos factores (factores de riesgo) desarrollen posteriormente una enfermedad determinada. Ahora bien,

el riesgo es una probabilidad, no una certeza, por lo que una asociación no implica necesariamente una relación de causa-efecto, como examinaremos con más detalle más adelante. Es decir, no todos los individuos expuestos al factor de riesgo desarrollarán la enfermedad, ni por el hecho de no estar asociados al efecto de tal factor de riesgo no la desarrollarán.

Antes de continuar con lo que será la descripción de las medidas de asociación o efecto, hemos de pararnos muy brevemente a detallar en qué consiste una tabla de contingencia o tabla de 2 x 2. Ya sea que se realice un ensayo clínico, un estudio de cohortes o un estudio de casos y controles, los resultados finalmente obtenidos se dispondrán en una tabla con cuatro casillas centrales: la del número de expuestos que presenta el acontecimiento en cuestión (A), la del número de expuestos que no presenta el acontecimiento en cuestión (B), la del número de no expuestos que presenta el acontecimiento (C) y la del número de no expuestos que no presenta el acontecimiento (D) (tabla 19).

Tabla 19. Tabla de contingencia o tabla de 2 x 2

	Individuos enfermos	**Individuos no enfermos**
Individuos expuestos	**A**	**C**
Individuos no expuestos	**B**	**D**

La incidencia de una enfermedad en una población se denomina riesgo absoluto. Este permite cuantificar la magnitud del riesgo de sufrir una enfermedad en un grupo de personas expuestas a un factor de riesgo. Por ejemplo, el riesgo absoluto de morir un año después de un infarto agudo de miocardio es la tasa de mortalidad calculada en un grupo de pacientes que han sufrido un infarto, alrededor del 8%.

Para determinar si existe asociación entre un factor de riesgo y una enfermedad, o bien si hay un aumento del riesgo en el grupo expuesto, debemos comparar el riesgo absoluto del grupo expuesto con el riesgo absoluto del grupo no expuesto. Esta comparación se realiza calculando el cociente riesgo absoluto de los expuestos/riesgo absoluto de los no expuestos. Si lo expresamos matemáticamente, entonces:

$$RR = \frac{\text{Riesgo del desenlace en presencia del factor de riesgo}}{\text{Riesgo del desenlace en ausencia del factor de riesgo}}$$

$$RR = \frac{(A)/A + B}{(C)/C + D}$$

- **Si RR = 1**, entonces no hay ningún tipo de asociación entre el factor de riesgo y la enfermedad; es decir, la exposición no aumenta ni disminuye el riesgo de padecer la enfermedad.

- **Si RR > 1**, significaría que el grupo expuesto tiene un riesgo absoluto mayor que el grupo no expuesto y, por tanto, la exposición aumenta el riesgo de padecer la enfermedad (la exposición es entonces un factor de riesgo).

- **Si 0 ≤ RR < 1** (el riesgo no puede tomar valores negativos), significa que el grupo expuesto tuvo un riesgo absoluto menor que el grupo no expuesto y, por tanto la exposición disminuye el riesgo de padecer la enfermedad (es un factor de protección).

El riesgo relativo, entonces, da magnitud a una probable asociación entre cierta exposición y una enfermedad. Cuanto más grande sea este valor más seguros podemos estar de esta relación, ya que no es lo mismo afirmar que el RR de padecer cáncer de pulmón debido al tabaquismo es de 1,35 (lo cual implicaría el 35% de aumento de riesgo por encima de los no expuestos) que un RR de 3,5 (lo cual implica un 250% más de riesgo).

En los estudios retrospectivos o de casos y controles, dado que los grupos están determinados a partir de la enfermedad, no es posible calcular tasas de incidencia y, por tanto, tampoco es posible calcular el riesgo relativo. En este tipo de estudios se utiliza una medida conocida como *odds ratio* (OR),

también conocida como *razón de productos cruzados* o *razón de ventajas*. Esta mide la ventaja de tener el factor de riesgo si la enfermedad está presente dividida entre la ventaja de tener el factor de riesgo si la enfermedad no está presente. La OR tiene una interpretación muy similar a la del riesgo relativo, es decir si la OR > 1, la exposición aumenta el riesgo; si OR < 1, la exposición confiere protección frente a la enfermedad, y si OR = 1, no hay asociación causal entre las variables. El cálculo de OR es el siguiente:

$$OR = \frac{A \times D}{B \times C}$$

En un estudio de casos y controles el investigador puede modificar a voluntad el número final de cada columna en la tabla de contingencia. Es decir, en este tipo de estudios el número de individuos que padecen y no padecen la enfermedad no refleja necesariamente la frecuencia natural de la enfermedad.

El riesgo relativo y la razón de productos cruzados son, por consiguiente, medidas fundamentales que empleamos para cuantificar la fuerza de una asociación entre un factor de riesgo y una enfermedad.

2.3.3. Selección de las pruebas estadísticas adecuadas

La investigación biomédica requiere del análisis estadístico, especialmente si existen características de interés que presentan variabilidad. Mediante la estadística seremos capaces de evaluar correctamente la evidencia obtenida a partir de un estudio experimental, y si bien la certeza absoluta no existe, la estadística es capaz de convertir en un valor algo que nos describa el grado de incertidumbre. Para cumplir con este objetivo, debemos seleccionar una prueba estadística apropiada y así responder a la cuestión en estudio[102].

El tipo de procedimiento estadístico que se utilice va a depender de: *a)* los datos que se analizan; *b)* la manera en que se distribuyen las variables, y *c)* de la pregunta de investigación que se intenta responder.

Antes de seleccionar un método, debemos tomar dos decisiones: elegir la *variable dependiente* y la *independiente* y qué tipo de datos constituyen cada una de esas variables. En los estudios de investigación, deseamos medir la relación entre una o más variables que en principio creemos que pueden ser predictoras –variables independientes– o determinantes de un cierto resultado o efecto –variable dependiente–.

El número de variables independientes determina el tipo de método estadístico que es apropiado para analizar los datos. Por ejemplo, si nos interesa el riesgo anual de asma bronquial en una comunidad infantil sin tener en cuenta la rinitis alérgica o cualquier otra característica de estos niños, aplicaríamos los métodos estadísticos de análisis univariantes. Estas técnicas se aplican a una serie de observaciones que contienen una variable dependiente y ninguna independiente. Para examinar el riesgo de asma bronquial en relación con el hecho de tener reflujo gastroesofágico, usaríamos los métodos de análisis bivariantes. Estos métodos se aplican a grupos de observaciones con una variable dependiente y una independiente. Por último, si nos interesara el riesgo de asma bronquial, en niños de diversas edades, sexo y afectados o no de rinitis alérgica, aplicaríamos los métodos de análisis multivariante. Estos métodos se utilizan para grupos de observaciones que consisten en una variable dependiente y más de una independiente, como la edad, el sexo y la rinitis.

Para seleccionar la prueba estadística, debemos determinar el tipo de datos que constituyen las mediciones de cada variable. Con el fin de categorizar los tipos de datos, realizaremos una primera distinción entre datos *continuos* y *discretos*. Cuando la variable se puede cuantificar solo en valores enteros como sería el número de hijos, el número de empleados de una fábrica, el sexo o el estadio de las enfermedades, se le denomina *variable cuantitativa discontinua* o discreta. Para cada una de estas variables podemos seleccionar dos valores entre los cuales no es posible imaginar otros. Por ejemplo, no podemos imaginar que el número de hijos de una familia se encuentre entre 1 y 2. Cuando se puede expresar en fracciones, como la edad, la concentración sérica de colesterol, la tensión arterial, la estatura o el peso, se les denomina variables *cuantitativas continuas*.

Los datos pueden definirse, además, por su escala de medida. Algunas mediciones discretas se realizan en una *escala ordinal*. Los datos en una escala ordinal tienen una ordenación o posición específica, como en el caso de los datos continuos, pero no es preciso que el intervalo entre mediciones consecutivas sea constante. Cuando hay determinado orden como clase social (alta, media, baja), orden de nacimiento (primero, segundo, tercero, etc.) se denominan *variables cualitativas ordinales*. Cuando no se puede aplicar algún tipo de ordenamiento a los datos discretos, decimos que se midieron en una escala nominal. Son ejemplos de características medidas con

datos discretos en escalas nominales el tratamiento, el sexo, la raza y el color de los ojos. Los datos que tratamos como nominales incluyen mediciones con dos categorías, aunque se pueda considerar que tienen un orden intrínseco, porque uno es claramente mejor que el otro (p. ej., vivo y muerto).

Así es que los pasos iniciales que se deben dar para seleccionar una prueba estadística son:

1. **Identificar una variable dependiente o respuesta** (desenlace, efecto, enfermedad) y todas las variables independientes o predictoras (factor de riesgo, exposición, tratamiento, diagnóstico) a partir de la pregunta que se intenta responder con el estudio.

2. **Definir la naturaleza de la variable dependiente de interés:** continua, ordinal o nominal. El orden en que se han enumerado indica la cantidad relativa de información que aporta cada uno de estos tres tipos de variables (continua > ordinal > nominal). Si la variable dependiente tiene un número ilimitado de valores uniformemente espaciados, se trata de una variable continua. Si la variable dependiente contiene un número de valores limitado que pueden seguir un orden, se trata de una variable ordinal. Una variable dependiente nominal simplemente identifica la presencia o la ausencia de una condición.

La naturaleza de la variable dependiente condicionará, por tanto, la prueba estadística que seleccionar. Como ya hemos dicho con anterioridad, no es ni mucho menos objetivo de este libro llevar a cabo una revisión exhaustiva de todas las técnicas de análisis estadístico. No obstante, nos detendremos en algunos conceptos que sí son importantes y que le pueden ayudar a interpretar el análisis de los resultados de cualquier artículo científico.

Técnicas aplicables a variables discretas: chi al cuadrado

Chi al cuadrado (χ^2) es un método estadístico usado para determinar si el resultado de un experimento puede ser causa del azar, y están involucradas variables discretas (o sea, del tipo todo o nada).

Técnicas de análisis estadístico aplicables a variables continuas

Descripción de una población: la distribución normal y el desvío estándar

Cuando deseamos *describir una población* con respecto a cierta *característica cuantitativa*, usamos la *media* o el *pro-*

medio como indicador de una *tendencia central* de los datos. Otros valores descriptivos son la *mediana* (valor por encima y por debajo del cual se encuentra el restante 50% de los datos) y la *moda* (valor que se da con mayor frecuencia en la población estudiada).

Sin embargo, la media de la muestra no deja de ser una estimación de la posición de la media poblacional. Para una descripción adecuada de la población se necesita de algún indicador de la variabilidad de los datos alrededor de este valor. El indicador más común es la desviación estándar (σ), que representa algo así como el promedio de las diferencias de los datos respecto a la media.

Muchas variables biológicas cuantitativas, como pueden ser los valores de laboratorio, la fracción de eyección ventricular, la tasa de filtrado glomerular, etc. poseen lo que se denomina una distribución normal al graficar la frecuencia relativa con que se encuentran. Esta gráfica, a la que ya nos referimos anteriormente en la figura 10, tiene forma de campana (curva de Gauss) y se denomina también *distribución gaussiana*, en la que la mayoría de los valores se encuentran alrededor de la media y la frecuencia de valores es progresivamente decreciente hacia los extremos. Para que un dato o valor pertenezca a una población de distribución normal, hablamos de los límites de confianza como los valores extremos de la media $\pm 2\sigma$, lo que comprende el 95% de la población normal. El intervalo de confianza es entonces el comprendido entre ambos límites. Por ejemplo, cuando un alergólogo compa-

ra las mediciones espirométricas con los valores previsibles para la talla, la edad y el sexo, es costumbre aceptar como normales los valores que caen dentro de cierto intervalo (p. ej., ± 2 desviaciones estándar a partir de la media previsible). La *capacidad vital forzada* (FVC por sus siglas en inglés) es el máximo volumen de aire espirado, con el máximo esfuerzo posible, partiendo de una inspiración máxima, que se considera normal cuando es mayor del 80% de su valor teórico. Es decir, una FVC del 80% de los valores previsibles o superior se considera normal. Cuando se encuentra en el 79-65% de los valores previsibles significa un trastorno leve; en el 64-50%, moderado; entre el 49% y el 35%, grave y < 35% muy grave.

Comparación entre grupos: test de la 't' de Student

Para comparar dos grupos de datos cuando la medida de interés es una variable continua, usamos el test de *t*. Supongamos que queremos comparar los valores de fracción de eyección del ventrículo izquierdo (FEVI) en pacientes que han recibido el fármaco A frente a los que han recibido el fármaco B. Para conocer si nuestros dos grupos son diferentes entre sí, lo que hacemos es ver si la variabilidad entre las dos medias grupales respecto a la media general es mayor que la variabilidad de las muestras respecto a la media de cada grupo. Para esto aplicamos el test de *t*.

Se calcula el valor de *t*, que es igual a la diferencia entre las medias de los dos grupos (dividida entre el error estándar). Luego se busca el valor de *t* en una tabla en la que está asig-

nada una probabilidad de que dicho valor corresponda a la hipótesis nula (o sea, no existe diferencia entre las medias). Si esta probabilidad (*p*) es pequeña (p. ej., < 0,05), la hipótesis nula es rechazada y aceptamos que ambos grupos de estudio son diferentes.

Comparación entre varios grupos: análisis de varianza (ANOVA)

La técnica apropiada para analizar variables continuas cuando existen más de dos grupos es el análisis de varianza, más conocido como ANOVA. Un ejemplo podría ser el valor de FEVI en tres grupos de pacientes tratados con tres fármacos diferentes.

Los principios del análisis ANOVA son los mismos que para el test de *t*. Bajo la hipótesis nula, la diferencia entre las medias de cada grupo y la media de la población general debería ser debida solamente al azar. Si demostramos que esto no es así, aceptaremos que las variaciones se deben a que, efectivamente, son grupos distintos, o sea que el efecto de las drogas es diferente. Aquí se debe calcular la varianza entre los grupos (intergrupal) y la varianza dentro de cada grupo (intragrupal), cuyo cociente da lugar al valor F. Este se busca en tablas y se observa el correspondiente valor de p, tal como se hace con el test de *t*.

Coeficiente de correlación (r)

Para saber la relación entre dos variables cuantitativas, calculamos el coeficiente de correlación, que es una estimación puntual de la fuerza de la asociación entre dos variables continuas. La medida de asociación entre los valores de la dosis de un biológico y la concentración de IgE en el paciente con dermatitis atópica se puede investigar mediante el coeficiente de correlación (r), número que oscila entre -1 y +1. La interpretación de los valores de r es la que sigue **(tabla 20)**:

Tabla 20. Valores de r

Valor de r	Interpretación
0,0-0,25	Correlación escasa/sin correlación
0,26-0,50	Correlación moderada
0,51-0,75	Correlación buena
> 0,75	Correlación excelente o muy buena

Un coeficiente r = -1 indica a su vez una perfecta relación inversa. Como en el ejemplo propuesto, el tratamiento de la dermatitis atópica con cualquier fármaco será tanto más eficaz cuanto mayor sea la bajada de la concentración de IgE circulante en plasma.

Regresión lineal

Muchas veces nos interesa utilizar la medida de una variable independiente continua para estimar la medida de una variable dependiente. Sigamos con nuestro ejemplo e imaginemos que queremos analizar la relación entre la dosis de un fármaco hipotético para el tratamiento del eczema en la dermatitis atópica. En concreto, deseamos estimar las actividades eczematosas que esperamos que estén asociadas en la población con diversas dosis del fármaco. Sin tener en cuenta si el muestreo fue aleatorio o intencionado, podemos establecer una ecuación lineal para estimar el valor medio de la variable dependiente (Y_i) para cada valor de la variable independiente (X_i). En nuestro ejemplo, la variable dependiente es la actividad eczematosa media y la variable independiente, la dosis del medicamento. La ecuación de una relación lineal en una población se describe mediante dos parámetros: una pendiente (β) y un punto de intersección ().

$$Y_i = \alpha + \beta X_i$$

El punto de intersección α estima la media de la variable dependiente cuando la variable independiente es igual a cero. Por tanto, el punto de intersección α de la ecuación lineal de la actividad eczematosa y la dosis estimaría la media del grado de eczema en los individuos que no han tomado el medicamen-

to. La pendiente de una ecuación lineal indica cuánto cambia la magnitud de la media de la variable dependiente por cada cambio de unidad en el valor numérico de la variable independiente. En nuestro caso, la pendiente de la ecuación que describe la extensión de eczemas en función de la dosis estima cuánto desciende la actividad eczematosa por cada unidad que aumenta la dosis del fármaco.

2.3.4. Significación estadística e importancia clínica: valoración de la causalidad

Consideraciones sobre el valor de 'p'

El concepto *significación estadística* se relaciona con la necesidad de probar hipótesis, situación a la cual los clínicos no suelen estar habituados y, quizás, sea una de las razones por las que confían tanto en el concepto de significación estadística y se dejan llevar por el valor de *p*. Sin embargo, las pruebas de significación estadística nos dicen muy poco de la magnitud de una diferencia o de la fuerza de una asociación[103]. Por ello, es necesario preguntarse no solo por el nivel de significación estadística ($p < 0{,}05$) de una asociación, sino también si es lo suficientemente grande para ser útil en el medio clínico.

Desde el punto de vista clínico, el concepto de significación estadística no es relevante, pues no se resuelve la incertidumbre. Se debe tener en cuenta que estamos hablando de un concepto matemático, por lo que una asociación estadísticamente significativa puede no ser clínicamente relevante, una asociación estadísticamente significativa puede no ser causal y una asociación estadísticamente no significativa puede deberse a un problema de tamaño de muestra insuficiente. Es decir, podemos encontrar asociaciones estadísticamente significativas y conceptualmente adulteradas. Por ello, hay que tener siempre presente que el término estadísticamente significativo no es garantía de calidad.

A la hora de interpretar el resultado de una prueba de significación estadística, hemos de tomar una serie de precauciones.

- **El rechazo de la hipótesis nula no sugiere causalidad.** Es un error grave, y lamentablemente muy frecuente en toda la literatura médica, asumir que una prueba estadísticamente significativa lleva asociada una relación de causa-efecto.

- **Un resultado no significativo no demuestra que la hipótesis nula sea cierta.** Un resultado no significativo solo indica que es compatible con la hipótesis nula porque la discrepancia es pequeña, por lo que no debemos pensar que el resultado negativo es sinónimo de hipótesis nula demostrada. Un resultado no significativo solo indica que los datos no consiguen aportar suficientes pruebas para dudar de la credibilidad de la hipótesis nula. De hecho, es perfectamente posible que la hipótesis nula sea falsa, pero no lo hemos detectado porque el tamaño de la muestra del estudio ha sido demasiado pequeña para descubrir el efecto esperado.

- **Un resultado estadísticamente significativo no tiene nada que ver con la clínica.** La expresión *muy significativo* es un término estadístico que indica la poca credibilidad de la hipótesis nula, y muy poco tiene que ver con la importancia clínica o biológica de la hipótesis.

- **La *p* no es una medida de la magnitud del efecto.** La expresión "muy significativo" no tiene nada que ver con la magnitud del efecto ni con la intensidad de relación entre variables. La significación estadística depende tanto de la magnitud del efecto investigado como del número de sujetos incluidos en el estudio, un valor que se convierte en una trampa muy habitual para intentar demostrar cualquier cosa[104]. Los estudios con un tamaño muestral muy pequeño tienden a resultar estadísticamente no

significativos, a pesar de que el efecto investigado tenga tamaño suficiente para ser considerado clínicamente interesante. Del mismo modo, estudios con muestras demasiado grandes tienden a dar resultados estadísticamente muy significativos, aunque el tamaño del efecto investigado sea irrelevante y carezca de interés clínico.

Por todo ello, a partir del artículo de Porta[105], he extraído una serie de recomendaciones que en el futuro podrán ir muy bien a cualquier alergólogo que quiera utilizar en sus estudios las pruebas de significación estadística **(tabla 21)**.

Tabla 21. Recomendaciones para tener en cuenta antes y después de hacer pruebas de significación estadística (modificada de Porta[105])

Antes	Después
Revisión crítica de la literatura científica	No nos dejemos fascinar por la *p*. Describir los resultados con independencia de la significación estadística. En los últimos años los trabajos son una explicación detallada de la aplicación de 20 supertest estadísticos impresionantes y la interpretación del autor sobre lo que quieren decir esos test, dejando al lector con la única opción que le queda: "Créetelo, lo dice la *p*"

Tabla 21. Recomendaciones para tener en cuenta antes y después de hacer pruebas de significación estadística (modificada de Porta[105]) *(continuación)*

Antes	Después
Elegir solo una o dos hipótesis operativas ("Quien mucho abarca...")	Calcular la magnitud de la asociación, la diferencia y el riesgo
Diseño del tipo de estudio de acuerdo con la pregunta de investigación (consultar a un bioestadístico). Un buen estudio con un mal análisis estadístico puede arreglarse. Un estudio mal diseñado de entrada no puede arreglarse con ninguna prueba estadística, por sofisticada que sea	Calcular el grado de significación *p* y los intervalos de confianza
Definición clara de las variables del estudio y de su relación (medirlo todo introducirá mucho ruido en el análisis)	Si $p > 0{,}05$, calcular la potencia estadística. Al menos podremos saber si necesitamos más casos
Selección de las pruebas estadísticas necesarias	Juicio de la relevancia clínica de los resultados, sean o no estadísticamente significativos
Decidir un riesgo de significación estadística (α) y un riesgo β y calcular el número de pacientes necesarios para poder extraer conclusiones	Discusión de los errores en el estudio y las explicaciones alternativas a los resultados obtenidos

(Continúa)

Tabla 21. Recomendaciones para tener en cuenta antes y después de hacer pruebas de significación estadística (modificada de Porta[105]) *(continuación)*

Antes	Después
Recolección solo de los datos esenciales y relevantes para poner a prueba nuestra hipótesis, asegurándonos de su veracidad	Propuesta de nuevos estudios para poder avanzar en nuestra investigación

Las pruebas estadísticas jamás sustituirán al juicio clínico. Es igual lo significativa que sea la *p*; lo importante es lo significativa que es la clínica.

La relevancia clínica

Por todo lo expuesto hasta ahora, es mucho más sensato hablar de relevancia clínica y no de significación estadística. 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 que se investiga, la vulnerabilidad, la morbimortalidad generada por el mismo, los costes y por su frecuencia entre otros elementos.

De este modo, las formas más apropiadas de representar los resultados de un ensayo clínico aleatorizado en investigación clínica son la *odds ratio*, el riesgo absoluto (RA), el riesgo relativo (RR), la reducción relativa del riesgo (RRR), la reducción absoluta del riesgo (RAR) y el número necesario de pacientes que tratar para reducir un evento (NNT). Este modo de presentar los resultados nos cuantifica el esfuerzo que hay que realizar para conseguir la reducción de un evento desfavorable. El presentar los resultados solo como reducción porcentual del riesgo relativo (RRR), aunque es técnicamente correcto, tiende a magnificar el efecto de la intervención al describir del mismo modo situaciones muy dispares. El cálculo del NNT representa el número de pacientes que tratar de manera experimental a fin de evitar que uno de ellos desarrolle un resultado negativo. Es por ello un estadístico excelente para determinar la significación clínica de un ensayo que, además, sea estadísticamente significativo. Cuanto más reducido es el NNT mayor es el efecto de la magnitud del tratamiento.

Valoración de la causalidad

A partir de todo lo expuesto hasta ahora, es mucho más sensato hablar de relevancia. ¿Puede el tabaco causar cáncer? ¿Puede el colesterol ser la causa de la enfermedad coronaria? ¿Y pueden los frutos secos causar trastornos alérgicos? El clínico tiene que hacer frente constantemente a controversias relacionadas con causas y efectos. Por esta razón, el lector de

la literatura médica debe comprender el concepto de causalidad manejado por los investigadores.

Un concepto clínico práctico de causalidad es el denominado causa contribuyente, que requiere de los siguientes criterios: *a)* la causa está asociada con la enfermedad –efecto– con más frecuencia que la esperada solo por azar; *b)* demostración de que la causa precede a la enfermedad, y *c)* demostración de que la modificación exclusiva de la causa altera la probabilidad del efecto. No requiere, sin embargo, que todos los sujetos que no presentan la causa contribuyente no presenten el efecto, ni que todos los que la posean desarrollen el efecto. En otras palabras, una causa clínicamente útil puede no ser ni necesaria ni suficiente, pero tiene que ser contribuyente. Su presencia debe aumentar la probabilidad de que se manifieste la enfermedad y su ausencia debe reducir dicha probabilidad. El proceso de análisis, que incluye pruebas de significación estadística y ajustes, ayuda a determinar si existe una asociación y si esta se debe a algún sesgo conocido. Sin embargo, para cumplir el segundo y el tercer criterios necesitamos basarnos en algo más que en análisis estadísticos[99].

El fenómeno de la causalidad requiere de un modelo para su descripción. Recurriremos al modelo determinista de Rothman, que define como *causa suficiente* un grupo de condiciones y acontecimientos mínimos que, inevitablemente inician o producen la enfermedad[107]. Dicho de otro modo, una causa es suficiente cuando la enfermedad es observada en los individuos

que presentan la causa. No obstante, la causa no se observa en todos los individuos con la enfermedad porque existen otras causas para ella. Para este modelo una causa se denomina *necesaria* cuando siempre debe anteceder una enfermedad. De acuerdo con este modelo, un factor es una causa de la enfermedad si la alteración de este resulta en una variación de la frecuencia del evento, de modo que al identificar y eliminar solo uno de los factores de la causa suficiente, se previenen los casos de la enfermedad que esta causa suficiente origina.

Al estudiar la causalidad se debe considerar que los factores de riesgo tienen diferentes roles en el desarrollo de la enfermedad. Se han determinado diferentes factores de causalidad:

- **Factores predisponentes** que crean un estado de sensibilidad hacia la enfermedad (edad, nivel educativo, ocurrencia previa de la enfermedad, etc.).
- **Factores facilitadores** de la manifestación de la enfermedad (hospitalizaciones, acceso a la atención médica, nutrición, etc.).
- **Factores desencadenantes** que están asociados con la aparición definitiva de la enfermedad (exposición a un medicamento, agentes tóxicos, traumatismo, etc.).
- **Factores potenciadores** que tienden a perpetuar o agravar la presencia de una enfermedad (exposición repetida a un agente tóxico, etc.).

Uno de los objetivos de la investigación biomédica es tratar de determinar si las asociaciones encontradas entre exposiciones y efectos son realmente de tipo causal, asegurándonos de que no existen sesgos, ni factores de confusión. Hay que valorar si la asociación detectada cumple una serie de consideraciones de causalidad **(tabla 22)**. Inicialmente descritas por el médico y estadístico Sir Austin Bradford Hill en 1965, cuantas más consideraciones se cumplan, más probable será que la asociación encontrada sea de tipo causal[108].

Cualquier investigación biomédica consiste en calcular el riesgo de la ocurrencia de una enfermedad con relación a uno o más factores de riesgo o exposiciones mediante los métodos estadísticos. En realidad, la estimación de riesgo representa la magnitud de la asociación entre el factor de riesgo y la enfermedad en estudio. Al examinar cualquier trabajo científico, hemos de ser capaces de dilucidar si la estimación de riesgos denota causalidad. Es decir, si lo que pretendemos estudiar tiene un efecto real sobre el desenlace de lo que se estudia.

Tabla 22. Consideraciones de Hill para establecer una relación causal (modificada de Hill[108])

Consideración	Comentario
Fuerza de una **asociación: las asociaciones más fuertes pueden ser** *causa*	Algunas asociaciones pueden tener variables de confusión que aumenten *falsamente* el efecto
Consistencia: **observación repetida de la asociación en poblaciones diferentes**	La consistencia no es esencial, ya que es posible que un factor de riesgo actúe en una comunidad pero no en otra. P. ej., una transfusión contaminada por un virus no tiene por qué causar infección a todos los transfundidos
Especificidad: **una causa conduce a un efecto único**	Un factor causal puede serlo para varias enfermedades, y una misma enfermedad puede tener diferentes causas
Temporalidad: **la causa debe preceder al efecto**	Es la consideración más fuerte y el único criterio posible de causalidad
Gradiente biológico: **presencia de la relación dosis-respuesta en la relación causa-efecto**	Una relación dosis-respuesta, aunque sea frecuente en medicina, no es necesaria para establecer la existencia de una relación de causa-efecto. P. ej., es posible que la radiación en dosis bajas y moderadas sea una causa contribuyente del cáncer de tiroides, mientras que las dosis altas maten realmente las células y, por ello, no contribuyan a su desarrollo

(Continúa)

Tabla 22. Consideraciones de Hill para establecer una relación causal (modificada de Hill[108]) *(continuación)*

Consideración	Comentario
Plausibilidad: **supone que comprendemos el proceso biológico que nos ocupa**	No siempre se dispone de información suficiente para definir la plausibilidad del proceso
Coherencia: **la relación causa-efecto se asemeja a lo que conocemos de la enfermedad**	Este criterio será inútil si se conoce poco de la enfermedad que pretendemos estudiar
Evidencia experimental: **evidencia de experimentos humanos y/o animales confirmando la asociación**	Es poco probable al trabajar con poblaciones humanas
Analogía: **relación de la causa-efecto con otras causa-efecto**	Solo limitada por la posible falta de imaginación del investigador

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado?

(1ª parte)

3

CONCLUSIONES

3. CONCLUSIONES

No quisiera terminar este primer libro sin destacar lo más importante de lo expuesto hasta ahora. Con este pequeño repaso, espero que seamos capaces de llevarnos una idea clara de lo que para el especialista médico ha de representar la lectura crítica.

Sobre la importancia de hacer una lectura crítica

- ✓ Leer con espíritu crítico ofrece la oportunidad de aumentar la efectividad de nuestra lectura, adquiriendo con prontitud las habilidades necesarias para excluir aquellos artículos que se juzguen de mala calidad y aceptar solo aquellos con la suficiente calidad metodológica.

- ✓ La medicina basada en la evidencia es el nuevo paradigma de la práctica clínica. Mediante un proceso sistemático de búsqueda, evaluación y aplicación de los hallazgos en investigación biomédica, el especialista podrá hacerse con una base esencial para la toma de decisiones en la

práctica clínica. En sus distintas fases (búsqueda de material bibliográfico, lectura crítica del mismo valorando su grado de evidencia e interpretación de los resultados) la medicina basada en la evidencia nos va a ayudar a distinguir lo que pudiera ser útil en la atención y tratamiento de nuestros pacientes.

✓ Los artículos científicos que caen en nuestras manos han de tener una estructura determinada: parte preliminar (título y resumen), introducción y objetivos, material y métodos, resultados, discusión y conclusiones. La parte preliminar debe "entrarnos por los ojos", es decir, ya el título debe orientarnos, en esencia, sobre la lectura de un artículo cuyo contenido será útil e interesante. Introducción y objetivos han de justificar la pregunta de la investigación. Material y métodos describen el modo de llevar a cabo el estudio, de manera que podamos reproducirlo con posterioridad. En los resultados, la exposición ha de ser objetiva, y es donde se sustentan las bases para el razonamiento empírico de la decisión clínica. Finalmente, discusión y conclusiones son la parte del estudio que debe convencer al lector de la fiabilidad y validez interna de los resultados, de su importancia y de su grado de utilidad clínica.

✓ Hoy, más que nunca, por la enorme cantidad y variedad de información a la que cualquier especialista puede tener acceso, resulta obligado deshacernos de algunos

mitos relativos a la publicación de un artículo científico. Mucho de lo que se publica carece de validez científica. El hecho de que el autor de un artículo goce de una vasta experiencia profesional o haya publicado otro artículo de gran impacto no es garantía en absoluto de que lo que estemos leyendo en el momento sea válido. En la literatura científica hay muchos tipos de sesgos, y siempre se tiende a publicar resultados positivos frente a los negativos.

- ✓ La lectura crítica nos ayudará a reducir la variabilidad impuesta por la práctica clínica que depende de la población, la oferta sanitaria del hospital en cuestión y de cada especialista.

- ✓ La lectura crítica ha de hacerse en tres etapas: selección de los artículos, análisis de la capacidad del artículo para contestar al problema que nos planteamos y la evaluación minuciosa de la metodología utilizada en aquellos que han superado las dos fases anteriores.

- ✓ Tanto las escalas de clasificación de la evidencia científica (clasificación revisada del Canadian Task Force on the Periodic Health Examination), como las guías CASP y CASPe son una herramienta muy útil de ayuda a los clínicos a la hora de decidir cuál es la nueva información que deben incorporar a su práctica cotidiana.

Criterios generales de la lectura crítica

- ✓ La investigación clínica va a ser de vital importancia para el especialista, al que le va a permitir renovar el conocimiento de la práctica médica y su posterior puesta en práctica. El clínico ha de saber así cómo hacer una valoración crítica de un artículo, juzgar si es válido, su grado de relevancia y, finalmente, la utilidad del mismo.

- ✓ Tras la selección de un artículo, lo primero será evaluar su validez interna y externa. La validez interna se refiere al grado en que los resultados del estudio son extraídos correctamente del mismo, mientras que la validez externa nos dice el grado en que estos resultados son generalizables a otras poblaciones distintas de la estudiada.

- ✓ Los clínicos han de ser consumidores inteligentes de bibliografía y, para ello, se han de conocer los criterios generales de la lectura crítica. La estrategia más utilizada sobre la evaluación de la validez consiste en responder a una serie de preguntas relacionadas con el diseño y el desarrollo de la investigación.

- ✓ ¿Cuál es la pregunta de investigación? Se trata de averiguar si el estudio responde o no a nuestra pregunta de investigación, es decir, a lo que pretendemos buscar (relevante, factible, novedoso, ético).

- ✓ ¿Se trata de un estudio original? Lo correcto sería preguntarnos si la investigación descrita nos aporta nuevos conocimientos al problema en estudio.

- ✓ ¿Por qué se hizo el estudio? Nos da la información necesaria para entender la pregunta principal del estudio, así como sobre cuán novedoso resulta y su importancia. La dirección del análisis va de lo general (antecedentes del problema) a lo particular (la pregunta, lo desconocido).

- ✓ ¿Qué tipo de estudio han desarrollado los investigadores? El mejor de los estudios para cada caso es el que presente menos problemas de validez y sea compatible con la disponibilidad de recursos. Los estudios clínicos pueden ser analíticos o descriptivos (finalidad del estudio), experimentales u observacionales (el control de la asignación de los factores del estudio) y longitudinales o transversales (la temporalidad del estudio).

- ✓ ¿Tiene el estudio un diseño adecuado a su objetivo? En esencia, cualquier trabajo de investigación médica ha de tener una buena pregunta a la que contestar, y para ello será necesario elegir el mejor diseño para hacerlo (comparativos, cruzados o por pares).

- ✓ ¿Se han evitado o minimizado los sesgos? Minimizar los sesgos en la medida de lo posible va a hacer que el especialista confíe más en los resultados expuestos en el artículo.

- ✓ ¿Es adecuada la población estudiada? El modo de incluir a los pacientes en los ensayos, el tipo de pacientes, la forma de muestreo y el tamaño de la muestra han de ser adecuados.

- ✓ ¿Dispone de una metodología adecuada? Hay que poner especial atención en el cálculo del tamaño de la muestra, el análisis según intención de tratar y la duración estimada del seguimiento y del número de pacientes que terminaron el estudio.

- ✓ ¿Y qué decir de la cuestión estadística? Es necesario saber si el estadístico aplicado a los resultados del estudio es el apropiado. Las mediciones de la investigación han de ser ajustadas de acuerdo con el método estadístico seleccionado y la posterior interpretación de los resultados que se expone debe ser la adecuada.

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado?

(1ª parte)

4

BIBLIOGRAFÍA

4. BIBLIOGRAFÍA

1. Sackett DL, Straus SE, Richardson WS, et al. Medicina basada en la evidencia. 1ª edición. Madrid: Elsevier España (ed); 2002.
2. Fernández E, García AM. Búsqueda y lectura crítica de artículos científicos. Barcelona: Signo; 2006.
3. Gómez M, Danglot C, Velásquez JL. Bases para la revisión crítica de artículos médicos. Rev Mex Pediatr. 2001;68:152-9.
4. Clava JJ, Ponce de León S, Vargas F. Cómo leer revistas médicas. Rev Invest Clin.1988;40:67-71.
5. Letelier LM, Manríquez JJ, Rada G. Revisiones sistemáticas y metaanálisis: ¿son la mejor evidencia? Rev Med Chile. 2005;133:246-9.
6. García Maldonado G, Sánchez-Juárez IG. Análisis crítico de la literatura científica: cómo analizar un artículo científico. Rev Med Salud Soc. 2011;1:1-15.
7. Aranda Torrelio E, Mitru Tejerina N, Costa Ardúz R. ABC de la redacción y publicación científica. 2a ed. La Paz: Elite; 2009.
8. Sackett D. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA. 1992;268:2420-5.
9. Guyatt GH. Evidence based medicine. ACPJ Club. 1991;112:A26.
10. Sackett DL, Rosenberg WMC, Gary JAM, et al. Evidence based medicine: what it is and what it isn´t. British Med J. 1996;312:71-2.

11. Sacket DL, Richardson WS, Rosenberg W, et al. Evidence-based medicine. New York: Churchill Livingstone; 1997.
12. González M et al. Medicina basada en la evidencia, medicina basada en la experiencia. Abril 2013. Disponible en: http://www.sefyp.es/artículos.
13. Day RA, Gastel B. How to write and publish a scientific paper. 7th ed. ABC-Clio, LLC, Santa Barbara; 2011.
14. Sackett D, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology: A Basic Science for Clinical Medicine. 2a ed. Little Brown & Co; 1991.
15. Dans AL, Dans LF, Guyatt GH, et al. Users' guides to the Medical Literature. How to decide on the applicability of clinical trial results to your patient. JAMA. 1998;279:545-9.
16. Martin Zurro A, Cano Pérez JF. Atención primaria. Conceptos, organización y práctica clínica. 5ª ed. Madrid: Elsevier; 2003.
17. Díaz Portillo J. Guía práctica de lectura crítica de artículos científicos originales en Ciencias de la Salud. Instituto Nacional de Gestión Sanitaria Ed. Madrid; 2010.
18. Guyatt G, Sackett DL, Cook DJ. Users' Guides to the Medical Literature II. How to Use an Article About Therapy or Prevention: B. What Were the Results and Will They Help Me in Caring for My Patients? JAMA 1994;271:59-63.
19. Bobenrieth MA. Lectura crítica de artículos originales en salud. Med Fam (And). 2001;2:81-90.
20. Cummings SR, Browner WS, Hulley SB, et al. Designing Clinical Research, 3rd ed., PA, Lippincott Willey & Wilkins; 2008.
21. Hoogenboom BJ, Manske RC. How to write a scientific article. Int J Sports Phys Ther. 2012;7:512-7.
22. Llorca J, Delgado M. Elaboración de proyectos de investigación sanitaria. 2ª ed. Bellaterra: Laboratori d'Estadística Aplicada i Modelització (UAB); 2009.
23. Laupacis A, Wells G, Richardson S, et al. Users' Guides to the Medical Literature: V. How to Use an Article About Prognosis. JAMA. 1994;272:234-7.

24. Vizoso FJ, Caballer J. La medicina basada en la evidencia. Cuad Psiquiatr Comunitaria 2005;5:25-33.
25. Levine M, Walter S, Lee H, et al. Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. JAMA. 1994;271:1615-9.
26. Jaeschke R, Guyatt G, Sackett DL. 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. 1994;271:389-91.
27. Biblioteca Facultad de Medicina. Universidad de Málaga. Normas de Vancouver: Guía básica. Cómo citar y elaborar referencias bibliográficas según las normas de Vancouver 2013. Documento disponible en: http://www.sld.cu/galerias.pdf.
28. Bohannon J. Who's afraid of peer review? Science. 2013;342:60-5.
29. Benach de Rovira J, Tapia JA. Mitos o realidades: a propósito de la publicación de artículos científicos. ACIMED. 1996;4:41-6.
30. Sokal A, Briamont J. Intelectual impostures. Profile Books, London; 1998.
31. Pessanha C. Criterios editoriales para la evaluación científica: notas para la discusión. ACIMED 2001;9 (especial).
32. Waldron T. Is duplicate publishing on the increase? British J Industr Med. 1992;304:1029.
33. Marušic A, Marušic M. Killing the messenger: should scientific journals be responsible for policing scientific brand? MJA. 2006;184:596-7.
34. Sterling RJ. Publication decision and their possible effects in inferences drawn from test of significance or vice-versa. J Am Stat Assoc. 1959;54:30-4.
35. Silva D, Llanes R, Rodríguez-Silva A. Manifestaciones impropias en la publicación científica. Rev Cubana Salud Pública. 2007;33:5-16.
36. Alcaide JF, Imaz I, González-Enríquez J. Búsqueda de evidencias. Una recopilación de recursos útiles en la evaluación de tecnologías sanitarias. Med Clin (Barc). 2000;114 (Suppl2):105-10.
37. Gómez de la Cámara A, Bravo R. Medicina basada en la evidencia. Cardiovasc Risk Factors 2001;10:349-75.

38. CASPe (Critical Appraisal Skills Programme Español). Disponible en: http://redcaspe/org.
39. Parrillas para la lectura crítica. 2012. Disponible en: http://redcaspe.org/drupal/?q=node/29.
40. Altman DG. Poor quality medical research: what can a journal do? JAMA. 2002;287:2765-7.
41. Sackett D, Straus S, Richardson W, et al. Medicina basada en la evidencia: cómo practicar y enseñar la EMB. 2ª ed. Madrid: Harcourt Eds; 2001.
42. Ridao-López M. Evaluación de la variabilidad geográfica en la práctica médica como herramienta de evaluación de prácticas sanitarias. Presupuesto y Gasto Público. 2012;68:193-210.
43. Wennberg L, Gittelson A. Variations in medical care among small areas. Sci Am. 1982;246:120-34.
44. The Dartmounth Atlas of Healthcare. Am Hosp Assoc Press, Chicago, Wennberg J, Cooper M. Eds.;1999.
45. Brook RH. Health services research: is it good for you and for me? Acad Med. 1989;64:124-30.
46. Moreno-Millán E. Variabilidad de la práctica clínica en la atención a urgencias. Emergencias 2007;19:222-4.
47. Sanz CL. Lectura crítica de artículos científicos. Cuad Psiquiatr Comunitaria. 2005;5:34-46.
48. Bobenrieth MA. Escritura y lectura crítica de artículos científicos. En: Burgos Rodríguez R. Metodología de investigación y escritura científica en clínica. Parte IV. Granada: Escuela Andaluza de Salud Pública; 1998.
49. Abad E, Monistrol O, Altarribas E, et al. Lectura crítica de la literatura científica. Enfermería Clin 2003;13:32-40.
50. Argimon JM, Jiménez J. Métodos de Investigación. Clínica y epidemiológica. 2.ª ed. Madrid: Harcourt Eds.; 2000.
51. Canadian Task Force on the Periodic Health Examination. The periodic health examination. Can Med Assoc J. 1979;121:1193-254.
52. Harris RP, Helfand M, Woolf SH, et al. for the Methods Work Group, Third U.S. Preventive Services Task Force. Current methods of the U.S.

Preventive Services Task Force: a review of the process. Am J Prev Med. 2001;20:21-35.
53. Jovell AJ, Navarro-Rubio MD. Evaluación de la evidencia científica. Med Clin (Barc). 1995;105:740-3.
54. Granados A. ¿Es útil la medicina basada en la evidencia para la toma de decisiones en sanidad? FMC. 1998;5:145-52.
55. Evidence-Based Medicine Working Group. La medicina basada en la evidencia. Guías del usuario de la literatura médica. JAMA (ed. esp.). 1997;24-108.
56. CASP and Health Care Libraries Unit. Evidence-based health care: an open learning resource for health care practitioners. Oxford: CASP and HCLU, 1999.
57. González de Dios J. Fundamentos de medicina basada en la evidencia. La valoración crítica de documentos científicos y su aplicabilidad a la práctica clínica: aspecto clave en la toma de decisiones basada en las mejores pruebas científicas. Evid Pediatr. 2006;2:37-40.
58. Viniegra VL. El camino de la crítica y la educación. Rev Inv Clin. 1996;48:139-58.
59. Espinosa AP, Viniegra VL. Lectura crítica en grupos escogidos de estudiantes de medicina. Rev Inv Clin. 1994;46:407-15.
60. Viniegra VL. La investigación como herramienta de aprendizaje. Rev Inv Clin. 1988;40:191-7.
61. Department of Clinical Epidemiology and Biostatistics. McMaster University Health Sciences Centre. How to read clinical journals: I. Why to read them and how to start reading them critically. Can Med Assoc J. 1981;124: 555-8.
62. Rico-Villademoros F, Hernando T. La introducción de un artículo científico original: el minusvalorado corazón del estudio. Panace@ 2011; 12:108-11.
63. World Medical Association: Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. Disponible en: http://www.wma.net/e/policy/b3.htm.

64. Franco-Monsreal J, Lara-Zaragoza EB, Villa-Ruano N, et al. Los estudios epidemiológicos. Temas de Ciencia y Tecnología. 2011;15:51-8.
65. Macmahon B, Pugh T. Principios y métodos de epidemiología. 2nd Ed., La Prensa Médica Mexicana, México; 1976.
66. Lilienfeld A, Lilienfeld D. Foundations of epidemiology. 2nd Ed., Oxford University Press, New York; 1980.
67. Hernández-Ávila F, Garrido-Latorre F, López-Moreno S. Diseño de estudios epidemiológicos. Rev Salud Publ Mex. 2000;42:144-55.
68. Kelsey JL, Thompson WD, Evans AS. Methods in Observational Epidemiology. New York: Oxford University Press; 1986.
69. Hennekens CH., Buring JE. Epidemiology in Medicine Boston: Litte, Brown and Company; 1987.
70. Chinn S. Longitudinal studies: objectives and ethical considerations. Rev Epidém Santé Publ 1989;37:417-29.
71. Rothman KJ. Modern epidemiology. Boston: Little-Brown; 1986.
72. Delgado-Rodríguez M, Llorca-Díaz J. Estudios longitudinales: concepto y particularidades. Rev Esp Salud Pública. 2004;78:141-8.
73. Lipnick RJ, Buring JE, Hennekens CH, et al. Oral contraceptives and breast cancer. A prospective cohort study. JAMA. 1986;255:58-61.
74. Rothman KJ, Greenland S. Modern epidemiology. 2nd Ed. East Washington Square, PA, Lippincott-Raven Publishers; 1998.
75. Arnedo-Pena A. Título en el manuscrito. Med Clin (Barc). 2004;123:490-5.
76. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Epidemiología clínica. Ciencia básica para la medicina clínica. 2ª ed. Madrid: Editorial Médica Panamericana; 1994.
77. Hill AB. A short textbook of medical statistics. Hodder & Stoughton, London; 1977.
78. Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials. 3rd ed. New York: Springer Verlag; 1998.
79. Buñuel-Álvarez J, Ruiz-Canela J. Cómo elaborar una pregunta clínica. Evid Pediatr. 2005;1:10-7.

80. Investigación y Ensayos Clínicos. En: Farmacia Hospitalaria. Tomo I. Publicación de la SEFH; 2002.
81. Sáenz-Campos D. Protocolos para investigación clínica: aspectos prácticos para su preparación. fármacos 2005;18:1-2.
82. Ávila HL. Introducción a la metodología de la investigación. Disponible en: http://www.cyta.com.ar
83. Picado C, Badiola C, Perulero N, et al. Validation of the Spanish version of the asthma control questionnaire. Clin Ther. 2008;30:1918-31.
84. Molina-Arias M. Lectura crítica en pequeñas dosis. Razones para dejar de leer un artículo. Rev Pediatr Aten Primaria. 2014;16:87-91.
85. Porta Serra M, Álvarez-Dardet C, Bolumar F, et al. La calidad de la información clínica: validez. Med Clin (Barc). 1987;89:741-7.
86. Greenhalgh T. How to read a paper: papers that summarize other papers. British Med J. 1997;315:672-75.
87. Hernández-Ávila M, Garrido-Latorre F, Salazar E. Sesgos en estudios epidemiológicos. Rev Salud Publ Mex. 2000;43:438-46.
88. Tejerizo L, Sánchez-Aparicio S, Castro-Aristi B. Bioestadística: el sesgo en la investigación (I). Tipos de sesgo. Toco-Gin. Pract. 1995;54:169-75.
89. Morabia A, Porta M. Causalidad y epidemiología. Investig Cienc. 2008;12:62-72.
90. Rebagliato M, Ruiz I, Arranz M, Metodología de investigación en epidemiología. Edit. Díaz de Santos, Madrid; 1996.
91. Robledo J. Población de estudio y muestreo en la investigación epidemiológica. Nure Investigación, n.º 10, noviembre 2004.
92. Robledo J. Diseños de muestreo probabilísticos (I). Nure Investigación, n.º 11, diciembre 2004.
93. Robledo J. Diseños de muestreo (II). Nure Investigación, n.º 12, febrero 2005.
94. Laporte JR. Principios básicos de investigación clínica. Fundació Institut Català de Farmacologia; 2007.
95. Essex-Sorlie D: Medical Biostatistics & Epidemiology. Appleton & Lange, Norwalk, USA; 1995.

96. McCormick PA, O'Keefe C. Improving prognosis following a first variceal haemorrhage over four decades. Gut. 2001;49:682-5.
97. Capurro D, Gabrielli N, Letelier LM. Importancia de la intención de tratar y el seguimiento en la validez interna de un estudio clínico randomizado. Rev Med Chile. 2004;132:1557-60.
98. Chen W, Nikolova D, Frederiksen SL, et al. Beta-blockers for cirrhotic patients with esophageal varices that have never bled (Protocol for a Cochrane Review). In: The Cochrane Library, Issue 3. Chichester, UK: John Wiley & Sons, Ltd.; 2004.
99. Riegelman RK, Hirsch. Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica. Biblioteca de la Organización Panamericana de la Salud. Little, Brown & Co. (Eds.) Boston, MA, USA; 1992.
100. Inferencia, estimación y contraste de hipótesis. Disponible en URL: http://www.ugr.es/~eues/webgrupo/Docencia/MonteroAlonso/estadisticaII/tema4.pdf
101. Díaz Portillo J. Guía Práctica del Curso de Bioestadística Aplicada a las Ciencias de la Salud Instituto Nacional de Gestión Sanitaria Ed. Madrid; 2011.
102. Núñez M. Evaluación de las técnicas diagnósticas: Análisis estadístico. Escuela Universitaria de Tecnología Médica Udela R, Montevideo, Uruguay. Comité de Tecnólogos de ALASBIMN; 2008.
103. Rebasa P. Entendiendo la "p < 0,001". Cir Esp. 2003;73:361-5.
104. Domènech JM. Comprobación de hipótesis. Pruebas de significación y pruebas de hipótesis. En: Métodos estadísticos en ciencias de la salud. Barcelona. Signo; 1998.
105. Porta M, Plasencia A, Sanz F. La calidad de la información clínica (y III): ¿estadísticamente significativo o clínicamente importante? Med Clin. 1998;90:463-8.
106. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. Br Med J. 1995;310:452-4.

107. Rothman K, Greenland S. Modern epidemiology, 3rd Edition. Philadelphia: Lippincott-Raven Publishers; 2008.
108. Hill A. The environment and disease: Association or causation? Proceed Roy Soc Med London. 1965;58:295-300.

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado?

(1ª parte)

5

NOTAS

NOTAS

NOTAS

NOTAS

NOTAS

NOTAS