

COMO LEER REVISTAS MEDICAS

I. Por qué leerlas y cómo empezar a leerlas en forma crítica*

DEPARTAMENTO DE EPIDEMIOLOGÍA CLÍNICA Y BIOESTADISTICA,
UNIVERSIDAD McMASTER, HAMILTON-ONTARIO, CANADA

Esta serie ríe. capítulos sobre Epidemiología Clínica Ka sida preparada para aquellos médicos que están rezagados en su lectura clínica. Probablemente esto incluya tanto a los que escribieron este capitulo como a sus lectores y esto es esperable si se considera que, para mantenerse actualizado con las 10 revistas más importantes en el área de la Medicina Interna, un médico debería leer 200 artículos y 70 editoriales al mes (1). En la actualidad se publican más de 20000 revistas biomédicas diferentes (más que las 14,000 de hace 10 años). Sólo para actualizarse en el tema de hepatitis viral se requeriría buscar entre 16,000 artículos publicados a este respecto, número que incluye únicamente a los publicados en el idioma inglés durante los últimos diez años.

La literatura biomédica se expande a un ritmo compuesto del 6 al 7% anual (2), y por lo tanto su número se duplica cada 10 a 15 años y se incrementa diez veces cada 35 a 50 años. Por el contrario, el tiempo disponible para que el médico lea la literatura clínica se ha ido recortando cada vez más por otras necesidades. Ante esta situación, las recomendaciones hechas en esta serie de capítulos enfatizaran la eficiencia, validez y aplicabilidad, y harán que el atareado lector clínico elimine un artículo poco después de iniciada su lectura, en lugar de perder el tiempo analizándolo para finalmente llegar, de todas maneras, a rechazarlo. Por ello, las guías propuestas permitirán al clínico separar en forma rápida el "trigo de la paja" de la literatura médica, de manera que

su destreza clínica mejore -y no empeore— con la información sobre innovaciones diagnósticas, pronósticas y terapéuticas.

El enfrentamiento del lector con un artículo dado puede ocurrir de dos maneras. Primero, el artículo llega a sus manos como resultado de su suscripción a la revista o porque alguien se lo proporcionó. Segundo, el lector puede encontrarlo durante la búsqueda de información que le ayudará a diagnosticar o a tratar a un paciente determinado. Ambas rulas reflejan la prioridad de mantenerse informado respecto a los avances médicos, prioridad que es pasada, por alto en la siguiente circunstancia:

En 1973, 230 hombres fueron identificados como hipertensos en una fundidora de acero localizado ni Hamilton, Ontario. Si las presiones diastólicas de estos sujetos se mantenían en 95 mm Hg o más durante 3 meses de observación, se les elaboraba una historia clínica completa y se les sometía a exámenes do laboratorio. Después de ello, cada paciente hipertenso, junto con su expediente, era referido a un médico, quien decidía si tratarlo o no y cómo hacerlo (3).

Dos terceras partes del grupo de hipertensos recibieron tratamiento y una tercera parte permaneció sin él. Un análisis retrospectivo subsecuente identificó tres determinantes importantes en la decisión clínica de tratar o no a un paciente. Las primeras dos determinantes fueron el nivel de presión diastólica (los hombres con una hipertensión más grave eran los que tenían más probabilidades de ser tratados) y la

Presencia de daño orgánico (aquellos con evidencia de complicación cardíaca, cerebral, renal, ocular o vascular eran los que tenían más probabilidades de ser tratados), ambas lógicas y apropiadas. Sin embargo, la tercera determinante en la prescripción de tratamiento fue el año de graduación del médico a quien era referido el paciente; los graduados en años mas recientes eran los que tendían a administrar tratamiento. En realidad, parece que los médicos, tanto viejos como jóvenes, estaban ejerciendo el tipo de medicina prevaleciente al momento de finalizar su entrenamiento. Se les había enseñado el manejo contemporáneo adecuado de la hipertensión, pero no se les había enseñando cuándo decidir cambiarlo.

Este ejemplo dramatiza uno de los retos continuos de la práctica médica: la necesidad (si en realidad se desea continuar haciendo más bien que mal a los pacientes) de reconocer y responder al imperativo de cambiar nuestras maniobras diagnósticas y terapéuticas de manera que se mantengan coherentes con innovaciones científicas válidas.

En este aspecto es fundamental y se extiende más allá del problema de la hipertensión. Por ejemplo, ¿debe continuar utilizándose el clofibrato en el tratamiento de hombres maduros asintomáticos que cursan con hipercolesterolemia moderada? (4) ¿Debe considerarse la terapia con ácido acetilsalicílico en hombres con isquemia cerebral transitoria? (5) Si un médico que atiende a estos pacientes aún no esté simplemente rezagado en su lectura, sin también fallando en su práctica clínica.

¿Cómo pueden los atareados clínicos enfrentarse a este reto de reconocer y responder a la necesidad de cambiar su práctica clínica? Las estrategias existentes incluyen educación continua, recertificación y similares. La estrategia que es el foco de atención de éste y de los capítulos subsecuentes es la lectura de revistas clínicas.

¿POR QUÉ LEEMOS REVISTAS CLÍNICAS?

Existen diversas razones por las que leemos revistas médicas. En la tabla 1 se listan 10 de ellas. La mayoría no son pertinentes al objetivo de esta discusión y serán analizadas someramente.

En primer lugar, por supuesto, se encuentra la lectura de revistas con el fin de impresionar a los demás. Los espectadores pueden ser gente no médica (pasajeros en un transporte público, parientes, etc.) o colegas clínicos (en la cafetería del hospital, durante el paso de visita, en las sesiones, etc.) Los primeros quedarán impresionados por nuestra erudición y los

Tabla 1
DIEZ RAZONES PARA LEER REVISTAS CLÍNICAS

1. Para impresionar a otros.
2. Para mantenerse actualizado sobre el medio profesional
3. Para comprender mecanismos pato-biológicos.
4. Para saber cómo maneja un clínico experimentado cierto problema.
5. Para saber si utilizar una prueba diagnóstica nueva o ya existente en el medio propio.*
6. Para conocer el cuadro clínico y la evolución de una enfermedad.*
7. Para determinar etiología o causalidad.*
8. Para distinguir el tratamiento útil de aquél que es inútil o aún dañino.
9. Para enfrentar demandas de información sobre necesidades, uso, calidad y costo-efectividad de la atención de la salud. **
10. Para ser provocado por las cartas al editor.

* Las razones marcadas con asterisco se cubrirán en detalle en esta serie de capítulos sobre Epidemiología Clínica.

** Una serie ulterior de capítulos sobre Epidemiología Clínica intentará desmistificar los artículos sobre estos aspectos.

segundos, por nuestra aparente habilidad para mantenernos al día. Segundo, leemos para estar al tanto de las noticias profesionales. Muchas revistas sirven como órganos de difusión de asociaciones médicas u otras sociedades científicas. A través de ellas podemos informarnos de las acciones de tales organizaciones así como de las de nuestros colegas.

Tercero, en ocasiones leemos revistas para entender mejor los avances, con frecuencia impresionantes, en la pato-biología de problemas clínicos que acostumbramos ver en nuestra práctica.

Cuarto, a menudo leemos para saber cómo un clínico experimentado, ampliamente respetado, estudia y trata una enfermedad determinada. Esto ocurre especialmente si la frecuencia con que vemos esa entidad no es suficiente como para que queramos decidir nosotros solos sobre el valor de tal o cual maniobra. Por supuesto, al permitir que una autoridad nos diga cómo manejar a nuestros pacientes, sacrificamos algo más que el juicio clínico. Generalmente, también renunciamos a la oportunidad de conocer si la evidencia clínica realmente es válida y aplicable en nuestra práctica.

Quinto, leemos revistas para averiguar si una cierta prueba diagnóstica, nueva o ya conocida, mejorará la exactitud, comodidad, seguridad o eficiencia con que estudiaremos a nuestros pacientes.

Sexto, leemos estas revistas para aprender más acerca del curso clínico y el pronóstico de las entidades que encontramos. Podría tratarse de algún problema ya bien conocido, cuyo curso y pronóstico en la actualidad está cambiando o volviéndose más claro. Alternativamente, podría ser una "nueva" entidad, co-

mo la legionelosis. Una vez terminada esta lectura, podremos decidir si es necesaria alguna intervención diagnóstica o terapéutica y así hacer un mejor papel tanto en reconfortar y aconsejar a los enfermos como en paliar sus malestares.

Séptimo, leemos revistas para aprender acerca de relaciones causales o etiológicas. Por un lado este aspecto nos sirve para poder orientar mejor a nuestros pacientes. Por ejemplo, si la presencia de atributos o hábitos como la obesidad, la falta de ejercicio y el estrés laboral, constituyen riesgos reales para la salud. Por otro lado, este conocimiento nos permite proteger a nuestros enfermos de los efectos adversos de medicamentos y de otras maniobras clínicas.

Octavo, leemos revistas para poder distinguir aquellas maniobras preventivas, terapéuticas y de rehabilitación que benefician realmente a los pacientes, de aquellas que simplemente significan una pérdida de tiempo y dinero para ellos (y para nosotros también) y que, de hecho, generan más mal que bien.

Noveno, leemos revistas para entender la "nueva ola" de proclamas, enjuiciamientos y amenazas acerca de las necesidades en salud, la calidad de la atención y la eficiencia de la atención clínica o de cualquier otro tipo.* Finalmente, leemos algunas revistas médicas (especialmente las británicas) para ser provocados por las cartas al editor. Después de que uno de los autores de esta serie fue llamado "bateo" en una carta al *Lancet* (6), está convencido de que el médico general británico ofendido no tiene igual para articular su ira. Las razones quinta a la octava de la tabla 1 resumen la esencia del sentir y responder a la necesidad de cambiar nuestra forma de abordar los problemas de diagnóstico, pronóstico, etiología y tratamiento. Ellas serán el foco de esta serie.

Las estrategias que estamos por sugerir asumen que los lectores clínicos ya están retrasados en su lectura y que, en adelante, nunca tendrán más tiempo para leer del que tienen ahora.

Por esta razón, y porque las guías que siguen invocan una más cercana atención a la sección de "Material y Métodos", así como a otros puntos que frecuentemente aparecen con tipo de tamaño más pequeño, muchas veces se recomendará deshacerse de un artículo por inútil, usualmente en base a evidencia bastante preliminar. Sólo a través del rechazo temprano de la *mayor parte* de los artículos, los atareados clínicos podrán concentrarse en los *pocos* que resulten ser válidos y aplicables a su propia práctica.

* Resulta tan grande el volumen y la importancia de esta nueva ola, que le dedicaremos una serie de Discusiones en Epidemiología Clínica a ella sola. Permanezca en sintonía.

LAS PRIMERAS CUATRO GUÍAS

La figura 1 es un flujo grama de cómo leer artículos en revistas clínicas y muestra que las cuatro guías iniciales, que se describen a continuación, son aplicables a todas las razones para que se lea un artículo.

Lea el título

¿Es el artículo potencialmente interesante o útil para su práctica? Si no es así, deséchelo y pase al siguiente artículo, a alguna otra tarea o vaya a la pista de *hockey*.

Revise la lista de autores

Además de ocasionalmente reconocer a un ex-compañero, el lector experimentado conocerá la trayectoria de muchos autores. Si esta trayectoria es de un trabajo cuidadoso, precavido y que se ha mantenido firme a la prueba del tiempo, lea el artículo. Si, al contrario, esta trayectoria consiste en una serie de conclusiones infundadas que permanecen de moda hasta que las cartas al editor las desmienten o revelan su constante perjuicio a la búsqueda de información sólida, deshágase del artículo. Sin embargo, muchos autores serán nuevos o desconocidos por lo que, al igual que el trabajo de los escultores no renombrados, el de estos autores merece al menos un vistazo a las siguientes consideraciones.

Lea el resumen

Aquí el objetivo es simplemente decidir si la conclusión, en caso de ser válida, será importante para usted como clínico. El punto no es si los resultados del artículo son verdaderos (puesto que difícilmente lo podrá hacer con leer solamente el resumen), sino determinar si los resultados, de ser válidos, le son útiles.

Aquellos quienes han estado leyendo revistas por algún tiempo recordarán que su formato ha cambiado sustancialmente a través de los años. A finales de los sesentas, la antigua sección de "Resumen y conclusiones", que solía estar al final de los artículos, se mudó al inicio del artículo y se convirtió en un resumen en revistas como: *Lancet*, *New England Journal of Medicine* y *British Medical Journal*; siguiendo el ejemplo del *Journal of the American Medical Association* y del *Canadian Medical Association Journal*. Este formato actualizado le ha hecho más fácil al atareado clínico, el estudio rápido de los artículos médicos.

* Obviamente, algunas veces uno no puede decir que los resultados son veraces aún después de haber leído todo el artículo.

Reflexione sobre el sitio

¿Es el lugar donde se realizó el estudio lo suficientemente similar al suyo, de manera que sus resultados, de ser válidos, sean aplicables a los pacientes de su consulta? Aquí hay dos puntos a considerar. Primero, ¿se cuenta con los recursos requeridos, o son suficientes la experiencia y la tecnología que están a su alcance, para permitirle poner en práctica la intervención descrita en el artículo? Segundo, ¿son los pacientes del lugar donde se originó el artículo similares a los suyos, con respecto a la gravedad de la enfermedad, el tratamiento, la edad, el sexo, la raza o en alguna otra característica clave que determine el evento clínico?

Puesto de otra manera, ¿son los resultados fácilmente transferibles a su propia práctica clínica? Por ejemplo, si usted es un médico familiar que ve pacientes a nivel primario, los resultados de estudios llevados a cabo en clínicas especializadas, en centros de referencia de tercer nivel, pueden no ser aplicables. Esto no es cuestión de esnobismo. Considere a la hipertensión: por años, los médicos de atención primaria han sido urgidos por los nefrólogos de centros de tercer nivel a llevar a cabo urografías excretoras cronometradas, exámenes endocrinológicos y otras extensas pruebas de laboratorio en pacientes hipertensos de reciente diagnóstico con el objeto de detectar aquellos que puedan ser curados (7). Una razón para esta recomendación fue el notable resultado de la hipertensión corregible quirúrgicamente, que se originó de las amplias investigaciones en pacientes hipertensos canalizados a centros de atención terciaria. Un ejemplo de esto se muestra en la tabla 2: a 6% de los hipertensos referidos a la Clínica Cleveland se les encontró una enfermedad curable quirúrgicamente (8); sin embargo, cuando una serie similar de estudios de laboratorio se llevó a cabo en pacientes hipertensos de la consulta general en Ontario, el hallazgo de hipertensión corregible por cirugía ocurrió en una proporción de enfermos 10 veces menor.

Ambas conclusiones son correctas. Su diferencia surge "de los "filtros" de referencia por los que los pacientes de la consulta general deben pasar antes de llegar a los centros de atención terciaria, como la Clínica Cleveland. Los médicos generales refieren, y los v centros de tercer nivel aceptan, de manera preferente a los pacientes con probabilidades relativamente altas de que la hipertensión sea secundaria. Así, los

* Se retomará esta consideración en capítulos ulteriores. Por ahora, una forma útil de considerar esta guía lleva a rephrasearla a su forma negativa: ¿Es el lugar tan diferente al suyo propio que sus resultados, aun cuando fuesen válidos, no fueran aplicable a él?

TABLA 2

EFFECTO DEL LUGAR DEL ESTUDIO SOBRE LA FRECUENCIA DE LA HIPERTENSIÓN CURABLE QUIRÚRGICAMENTE

| Causa de la hipertensión | % de pacientes | |
|--|---|--|
| | Canalizados a la clínica Cleveland (n = 4,939) ⁸ | En la consulta general en Ontario (n = 665) ⁹ |
| Feocromocitoma | 0.2 | 0 |
| Síndrome de Cushing | 0.3 | 0.2 |
| Aldostcronismo primario | 0.4 | 0 |
| Coartación de la aorta | 0.6 | 0.2 |
| Alteración rnovascular | 4.4 | 0.2 |
| Alteración del parenqui- ma renal | 5.2 | 4.7 |
| Uso de anticonceptivos orales | ---- | 0.2 |
| Poliomielitis | 0.0 | 0.2 |
| Hipertensión esencial | 88.9 | 94.3 |
| Hipertensión susceptible de manejo quirúrgico | 6.0 | 0.6 |

pacientes son soplos subcostales o con concentraciones bajas de potasio sérico son referidos a los centros de atención terciaria, lo que ha determinado la alta frecuencia de hipertensiones curables, quedándose en casa aquellos pacientes sin estas características. Es debido a estos filtros de referencia que, en los centros de atención primaria, deben realizarse diferentes pruebas diagnósticas que en los de atención terciaria. Solamente a través de poner atención al lugar donde se llevó a cabo el estudio reportado, es como reduciremos la aplicación errónea de sus resultados.

LA RAMIFICACIÓN DE LOS CAMINOS

Así, debemos ver de manera crítica a la experiencia publicada y no dar como ciertas a las "Conclusiones" ¹ sin considerar los "Materiales y Métodos" (10).

Aún el revisor y las políticas editoriales de las mejores y más respetables revistas no están exentos de error y una sola suscripción puede proveernos tanto de la verdad como de toda una gama de sesgos. De esta manera, desconocemos otra alternativa para los lectores clínicos (una vez que estén satisfechos con el título, los autores, el resumen y el lugar) que no sea el invertir tiempo en revisar la sección de Métodos del artículo y así evitarse perdidas de tiempo y dinero más tarde, al ejecutar procedimientos clínicos inútiles e incluso dañinos.

Durante el curso de los próximos artículos veremos que el clínico presuroso que acepta las conclusiones de un artículo después de sólo leer el resumen, lo hace poniendo en riesgo considerable a la verdad y los pacientes. Así, aunque en ocasiones con sólo leer el resumen del artículo pueda darse cuenta que carece

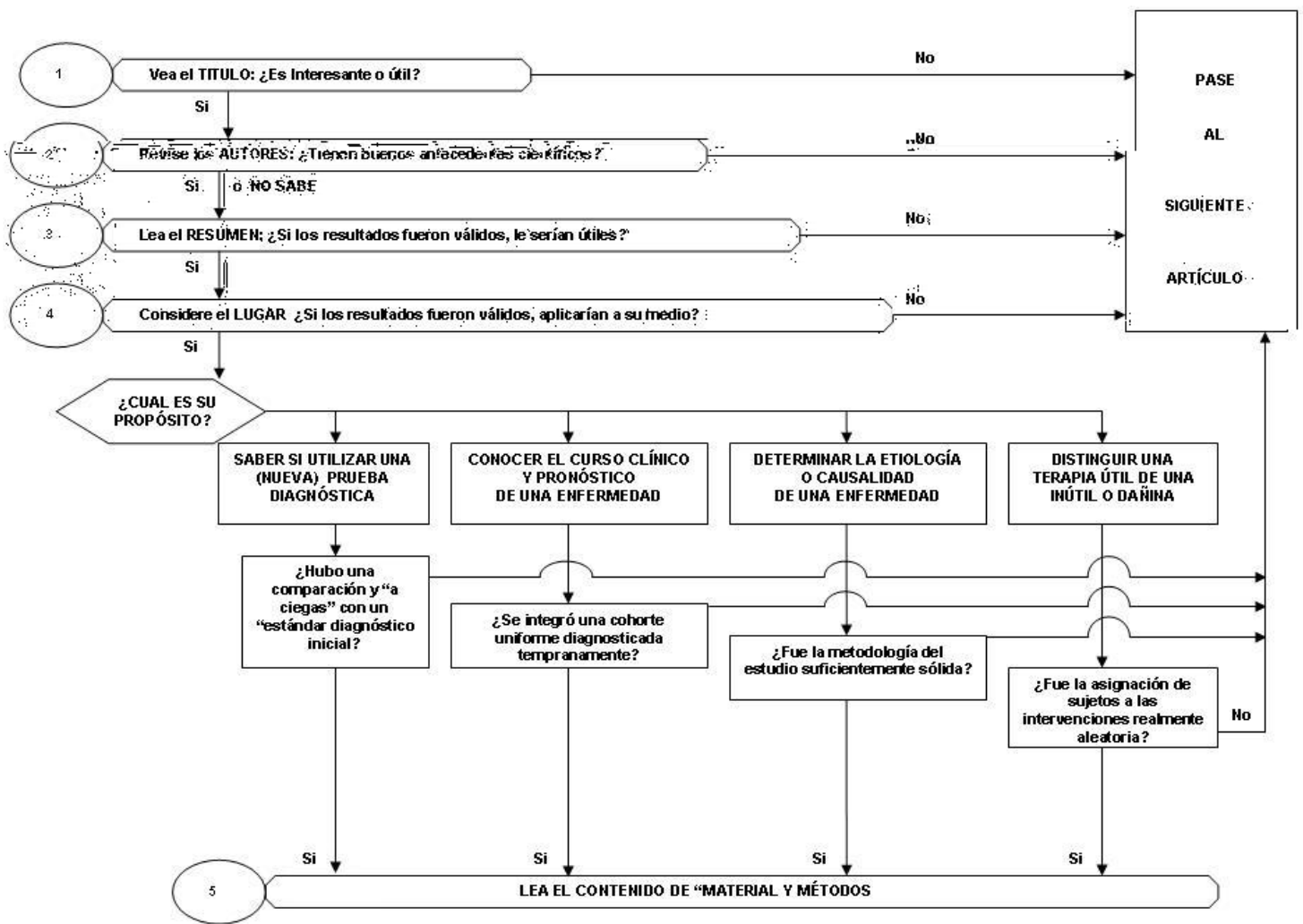


FIG. 1. Los primeros pasos de cómo leer artículos en una revista clínica

de validez (por ejemplo, cuando la eficacia de una intervención riesgosa se ha sustentado en el testimonio de los sobrevivientes) tal tipo de inspección casi nunca puede decirnos si un artículo es *válido*. Este juicio crucial nos lleva a tener esa combinación particular de sano escepticismo y sentido común práctico al que Ernest Hemingway llamaba "detección de inmundicias", al examinar la sección de Métodos del artículo. En este momento, las guías se ramifican dependiendo de la intención del lector, como se muestra en la figura 1. El siguiente capítulo de esta serie sobre Epidemiología Clínica tratará sobre las guías a usarse cuando se lea una revista clínica para decidir si usar una prueba diagnóstica específica.

REFERENCIAS

1. Warren KS (ed) : Qualitative aspects of the biomedical literature. En: Coping with the Biomedical Literature, New York (en prensa).
2. Price DS: The development and structure of the bio-medical literature. Ibid.
3. Sackett DL, et al: Randomised clinical trial for improving medication compliance in primary hypertension. *Lancet*, 1: 1205, 1975.
4. Oliver MF, et al: A co-operative trial in the primary prevention of ischaemic heart disease using clofibrate. Informe del Committee of Principal Investigators. *Dr Heart J*, 40: 1069, 1978.
5. Canadian Cooperative Study Group: A randomized trial of aspirin and sulfinpyrazone in threatened stroke. *N Engl J Med*, 299: 53, 1978.
6. Hart JT: Screening for hypertension (C). *Lancet*, 2: 1300, 1974.
7. Melby JC: Extensive hypertensive work-up: pro. *JAMA*, 231: 399, 1975.
8. Gifford RW Jr: Evaluation of the hypertensive patient with emphasis on detecting curable causes. *Milbank Mem Fund Q*, 47 (pt 2): 170, 1969.
9. Rudnick KV, et al: Hypertension in a family practice. *Can Med Assoc J*, 117: 492, 1977.
10. Spodick DH: On experts and expertise: the effect of variability in observer performance. *Am J Cardiol*, 36: 529, 1975.

COMO LEER REVISTAS MEDICAS

II. Para aprender sobre una prueba diagnóstica*

DEPARTAMENTO DE EPIDEMIOLOGÍA CLÍNICA Y BIOESTADISTICA,
UNIVERSIDAD McMASTER, HAMILTON-ONTARIO, CANADA

En el primer capítulo de esta serie (Rev. Inv. Clin. Mex. 1988); se presentaron 10 razones para leer revistas clínicas y un diagrama de flujo con guías para leerlas (figura 1). Este diagrama sugiere cuatro guías generales aplicables a cualquier artículo (referentes al título, los autores, el resumen y el lugar de elaboración) y enfática que las guías subsecuentes para leer o descartar una publicación dependerán de las razones por las cuales se lee esta.

A continuación se presentan las guías para leer artículos que describen pruebas diagnósticas, ya conocidas o nuevas.

Primero es necesario proporcionar algunas definiciones.

El nivel sérico de tiroxina (T_4) puede ser medido en, por lo menos, cuatro circunstancias diferentes. En la primera de ellas, los clientes que acuden a un centro comercial o los miembros de un club de ancianos son invitados a someterse a una determinación gratuita de T_4 ; este estudio de voluntarios aparentemente sanos procedentes de la población general, que busca separarlos en grupos con probabilidades altas y bajas de enfermedad tiroidea, se denomina *tamizaje* o *cribaje*. En la segunda circunstancia, los pacientes que asisten al consultorio de un médico por una enfermedad cualquiera, pueden tener incluida en forma rutinaria una determinación de T_4 a la serie de exámenes solicitados con el fin de diagnosticar la causa de sus molestias; este estudio de pacientes, que está orientado a la búsqueda de enfermedades ajenas a la razón de consulta, recibe el nombre de *detección de casos*. En la tercera circunstancia, una determinación

de T_4 puede ser solicitada para explicar la causa exacta de la enfermedad del paciente; esto se conoce como *diagnóstico*. Finalmente, una determinación de T_4 puede ser solicitada en un paciente: que recibe suplemento hormonal o que ha recibido yodo radioactivo, a fin de estudiar el *efecto del tratamiento*.

Este capítulo se orientará principalmente a la tercera de las circunstancias descritas: el diagnóstico. En los capítulos subsecuentes se hará mención de las circunstancias restantes.

GUIAS PARA LEER ARTÍCULOS SOBRE PRUEBAS DIAGNOSTICAS

¿Qué debe buscarse en la sección de "Material y Métodos" cuando se tiene un artículo que parece describir una prueba diagnóstica útil (esto es, que tiene un título interesante, los autores tienen buenos antecedentes científicos, el resumen refiere que sería muy útil si realmente funciona como se dice y el lugar de desarrollo es similar al suyo) ?

Los ocho elementos necesarios para hacer una evaluación clínica adecuada de una prueba diagnóstica aparecen en la tabla 1 (1-4). Estos son a su vez guías para el lector clínico y serán, considerados en orden.

1. ¿Se hizo una comparación independiente y "a ciegas" con un "estándar diagnóstico ideal"?

Los pacientes que tienen la enfermedad de interés (diagnosticada por medio de un "estándar" diagnóstico ideal", v. gr.: biopsia) y aquéllos que no la tienen (en base al mismo estándar ideal) debieron ser sometidos a la prueba diagnóstica en estudio, y ésta debió ser interpretada por mé-

Solicitud de sobretiros: Unidad de Epidemiología Clínica, c/o Dr. Samuel Ponce de León, Instituto Nacional de la Nutrición Salvador Zubirán, Vasco de Quiroga 15, Delegación Tlalpan, 14000 - México, D. F. México.

* Traducción del original publicado en el Canadian Medical Association Journal, vol. 124; pp. 703-710, 1981.

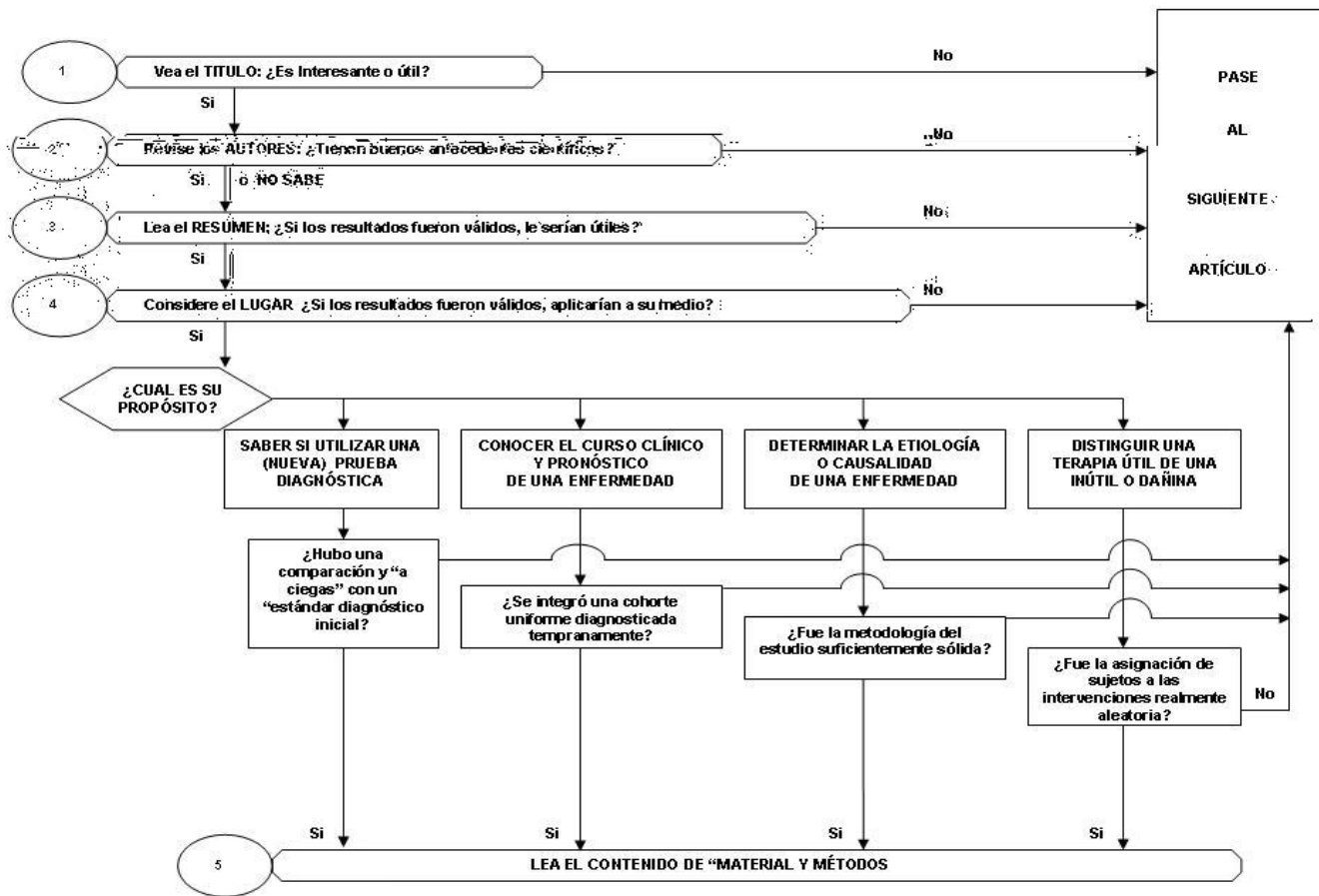


FIG. 1. Los primeros pasos de cómo leer artículos en una revista clínica

TABLA I

ELEMENTOS PARA UNA EVALUACIÓN CLÍNICA
ADECUADA DE UNA PRUEBA DIAGNÓSTICA

1. ¿Se hizo una comparación independiente y "a ciegas" con un "estándar diagnóstico ideal"?
2. ¿Se incluyó un espectro adecuado de pacientes, es decir, de sujetos con enfermedad leve y grave, tratada, y no tratada, e individuos con enfermedades diferentes pero cuadro clínico similar?
3. ¿Se describió en forma adecuada el medio donde se desarrolló el estudio al igual que el filtro por el que pasaron los pacientes?
4. ¿Se determinó la reproducibilidad de los resultados de la prueba (precisión) y de su interpretación (variación de observador)?
5. ¿Se definió en forma adecuada el término "normal"?
6. Si la prueba se propone como parte de un grupo o secuencia de pruebas, ¿se determinó su contribución a la validez global?
7. ¿Se describieron con suficiente detalle las tácticas para desarrollar la prueba, de manera, que ésta pueda replicarse con exactitud?
8. ¿Se determinó la "utilidad" de la prueba?

dicos que no sabían (esto es, que estaban "a ciegas") si un paciente dado realmente tenía o no la enfermedad. Después, los resultados de la prueba diagnóstica debieron compararse con el estándar ideal.

El método más directo de ilustrar la comparación de una prueba diagnóstica con el estándar ideal es el de una tabla de "dos por dos" o "cuádruple" (tabla 2). Las palabras clave en tales comparaciones son *sensibilidad*, *especificidad* y *valor predictivo*. Si el lector no encuentra por lo menos las dos primeras palabras en el resumen, debe irse con cuidado. Si no puede encontrar, o no puede construir, una tabla "cuádruple" después de una rápida revisión de la sección de resultados, es muy probable que no valga la pena continuar leyendo; en tal caso conviene eliminar el artículo y continuar con el siguiente.

Si el artículo sobrevive a esta rápida prueba de tamizaje, la comparación de los resultados de la prueba diagnóstica con los del estándar ideal habrá de brindar al lector una información muy útil. He aquí los conceptos básicos:

Primero, el estándar ideal se refiere al diagnóstico definitivo obtenido por biopsia, cirugía, autopsia, se-

TABLA 2
TABLA DE 2 X 2 QUE DEMUESTRA UNA COMPARACIÓN "A CIEGAS" CON EL
"ESTÁNDAR IDEAL"

| | | Estándar ideal | | |
|--|---|----------------------------------|------------------------------------|----------------------|
| | | El paciente tiene la enfermedad | El paciente no tiene la enfermedad | |
| Resultado de la prueba (conclusión basada en los resultados de la prueba) | Positivo: El paciente parece tener la enfermedad | Verdaderos positivos a | Falsos positivos b | a + b |
| | Negativo: El paciente parece no tener la enfermedad | Falsos negativos c | Verdaderos negativos d | c + d |
| | | a + c | b + d | a + b + c + d |

Propiedades estables:

$$a/(a + c) = \text{sensibilidad}$$

$$d/(b + d) = \text{especificidad}$$

Propiedades dependientes de la frecuencia:

$$a/(a+b) = \text{valor predictivo positivo}^*$$

$$d/(c+d) = \text{valor predictivo negativo}$$

$$(a + d)/(a + b + c + d) = \text{exactitud}$$

$$(a + c)/(a + b + c + d) = \text{prevalencia}$$

* El valor predictivo positivo puede también ser calculado de otras maneras. Una de ellas recurre al teorema de Bayes:
(prevalencia) (sensibilidad)

$$(\text{prevalencia}) (\text{sensibilidad}) + (1 - \text{prevalencia}) (1 - \text{especificidad})$$

guimiento a largo plazo, o por otro estándar reconocido. Si el lector no puede aceptar el estándar ideal (razonablemente, ya que ¡nada es perfecto!), entonces conviene que elimine el artículo.* Si acepta el estándar ideal, debe considerar entonces la prueba diagnóstica: ¿tiene ésta algo que ofrecer que el estándar ideal no tenga? por ejemplo, ¿implica menos riesgo?, ¿es menos molesta para el paciente?, ¿es menos costosa o. puede aplicarse en etapas más tempranas de la enfermedad? Nuevamente, si la prueba diagnóstica propuesta no ofrece ventaja teórica alguna sobre el estándar ideal, ¿para qué seguir leyendo?

Habiendo concluido que vale la pena continuar, el lector está preparado para analizar la comparación entre los resultados de la prueba diagnóstica y los del estándar ideal. Existen algunos elementos útiles de esta comparación, que habrán de analizarse por separado, introduciéndose al mismo tiempo la terminología técnica asociada.

Los primeros dos elementos consideran qué tan correctamente identifica la prueba diagnóstica a los pacientes con y sin la enfermedad de interés. Considérense las columnas de la tabla 2. El estándar ideal ha identificado (a+c) pacientes con la enfermedad de interés, y "a" pacientes tuvieron resultados positivos en la prueba

diagnóstica. De aquí que un índice de la capacidad de la prueba para detectar la enfermedad cuando está presente es $a/(a+c)$; usualmente, este índice se presenta como porcentaje y, para una mejor comunicación, se conoce como *sensibilidad*. En forma similar, la capacidad de la prueba diagnóstica de identificar correctamente la ausencia de la enfermedad se expresa en términos de la siguiente columna de la tabla, es decir, como $d/(b+d)$; este índice se conoce como *especificidad*. La sensibilidad y la especificidad pueden ser consideradas como propiedades estables de la prueba, dado que no cambian cuando se estudian diferentes proporciones de pacientes enfermos y sanos; éste es un aspecto importante, al cual se regresará más adelante.

Considérese la situación clínica usual. Cuando el clínico intenta diagnosticar la enfermedad de un paciente, no cuenta con los resultados de un estándar ideal para comparar. (Si contara con ellos, no se preocuparía de solicitar una prueba diagnóstica menos definitiva, dado que ya contaría de antemano con más información que la que ésta pudiera proporcionar). Nótese que en este caso se está operando horizontalmente en la tabla 2, no verticalmente. De aquí que, al juzgar el valor de una prueba diagnóstica, lo que realmente se desea conocer no es su sensibilidad y especificidad, sino el significado de sus resultados cuando éstos son positivos o negativos. Esto es, se quiere conocer qué tan bien predecirán los resultados

* El estándar ideal no debe incluir los resultados de la prueba diagnóstica evaluada, dado que el "sesgo de incorporación" resultante invalidaría toda la comparación (3).

de la prueba a los del estándar ideal. Si esta predicción es lo suficientemente buena, la prueba diagnóstica propuesta se verá integrada al paquete de recursos diagnósticos del lector.

El interés se centra entonces en las propiedades horizontales de la prueba, diagnóstica. Entre los pacientes con resultado positivo ($a+b$), ¿qué proporción recibe el diagnóstico correcto? Esta proporción, $a/(a+b)$, expresada en porcentaje, recibe el nombre de *valor predictivo positivo*.

En forma similar, la proporción $d/(c+d)$ expresa qué tan bien predice un resultado negativo de la prueba; la ausencia de la enfermedad en cuestión; esta proporción se conoce como *valor predictivo negativo*.

Otra propiedad de interés es la tasa global de concordancia entre la prueba diagnóstica y el estándar ideal. La tabla 2 muestra que ésta puede ser expresada como la fracción $(a + d)/(a+b+c+d)$, y usualmente recibe el nombre de *exactitud**.

Si el valor predictivo de una prueba diagnóstica es el aspecto de mayor interés clínico, ¿para qué perder el tiempo considerando su sensibilidad y especificidad? La razón es fundamental y tiene importantes implicaciones, no sólo para el uso racional de las pruebas diagnósticas, sino también para la educación básica de los clínicos. Los valores predictivos positivo y negativo de una prueba diagnóstica varían ampliamente, dependiendo de la proporción de individuos enfermos dentro del grupo de pacientes sometidos a la prueba -en la tabla 2 esto corresponde a la proporción $(a+c)/(a+b + c+d)$ y recibe el nombre de *prevalencia*.

Aunque la sensibilidad y la especificidad de una prueba diagnóstica permanecen constantes (o "estables"), ante cambios en la prevalencia de una enfermedad, los valores predictivos y la exactitud pueden variar en forma marcada (y por ello son "inestables"). Este no es un problema teórico, 3ado que en la vida real la prevalencia de una misma condición varía en forma considerable en poblaciones provenientes de diferentes niveles de atención, tal como se observó en el ejemplo de la hipertensión arterial analizado en el capítulo precedente. Más aún, considérese que durante el desarrollo inicial de la mayoría de las pruebas diagnósticas, éstas son evaluadas en números similares de sujetos con y sin la enfermedad de interés (esto es, una prevalencia fija de 50%), que es, casi siempre, una prevalencia mayor a la observada en la práctica clínica, aún en centros de atención de tercer nivel.

* Galen y Gambino (4), quienes escribieron un libro muy comprensible sobre este aspecto, denominaron a esta propiedad "eficiencia". Nosotros no lo haremos así.

Tomando en cuenta que para el uso racional de las pruebas diagnósticas es fundamental la comprensión del efecto que tiene la prevalencia sobre sus propiedades estables e inestables, y considerando que la resolución activa de problemas es superior a la asimilación pasiva, se invita al lector a resolver el siguiente problema (1).

Algunos investigadores estudiaron cuidadosamente un grupo de hombres referidos por dolor precordial. Después de someterlos a una prueba de esfuerzo (la prueba diagnóstica) y a una arteriografía coronaria selectiva (el estándar ideal) se obtuvieron los resultados que aparecen en la tabla 3 (5). La capacidad de la prueba de esfuerzo para predecir los resultados de la arteriografía coronaria selectiva se reveló en su valor predictivo positivo de 89% (el porcentaje de hombres con prueba de esfuerzo positiva cuya arteriografía demostró una estenosis del 75% o mayor) y en su valor predictivo negativo de 63% (el porcentaje de hombres con prueba de esfuerzo negativa cuya arteriografía mostró una estenosis menor a 75%). En base a esto, los autores concluyeron: "la positividad de la prueba de esfuerzo en sujetos masculinos es útil para predecir la presencia de enfermedad coronaria significativa; sin embargo, un resultado negativo no es confiable para descartar la presencia de enfermedad significativa" (5).

Como se desprende de los resultados arteriográficos (estándar ideal): $(a+c) / (a+b+c+d)$, o $104/195$, ó 53% de los pacientes, tuvieron estenosis coronaria marcada -un grupo de pacientes altamente seleccionado. ¿Qué ocurriría si los entusiastas adoptaran la prueba de esfuerzo en forma amplia con objeto de detectar enfermedad coronaria significativa en sujetos masculinos que desean correr o desarrollar otras actividades deportivas independientemente de la presencia o ausencia de dolor precordial? En tal caso, ¿sería útil una prueba de esfuerzo positiva?

Los resultados de la aplicación de esta prueba a un grupo menos seleccionado de hombres son completamente predecibles (tabla 4). Si la prevalencia real de estenosis coronaria marcada (evaluada por la arteriografía como estándar ideal) fuese sólo de $1/6$ ($104/624$ ó 17%) en lugar de $1/2$ ($104/195$ ó 53%), el valor predictivo positivo de la prueba se reduciría de 89% a 57% y su valor predictivo negativo se incrementaría de 63% a 91% lo opuesto a la situación original.**

* Los autores del trabajo referido en este ejemplo no hicieron tal recomendación.

** Este caso hipotético se aproxima a lo que realmente aconteció en el grupo de mujeres del trabajo referido (5). Gruesamente, una sexta parte tuvo una estenosis del 75% o mayor y la prueba de esfuerzo tuvo una sensibilidad del 50% y especificidad del 78% (valores cercanos a los observados en hombres), y valores predictivo positivo y negativo

TABLA 3

PRUEBA DE ESFUERZO COMO PREDICTOR DE ESTENOSIS CORONARIA EN PRESENCIA DE UNA PREVALENCIA DEL 50% (5).

| | | > 75% de estenosis | | |
|--------------------|----------|--------------------|---------|-----|
| | | Presente | Ausente | |
| Prueba de esfuerzo | Positiva | 55 | 7 | 62 |
| | | a | b | |
| | Negativa | 49 | 84 | 133 |
| | | 104 | 91 | 195 |

Valor predictivo positivo = $a/(a + b) = 55/62 = 89\%$

Valor predictivo negativo = $d/(c + d) = 84/133 = 63\%$

Sensibilidad = $a/(a + c) = 55/104 = 53\%$

Especificidad = $d/(b + d) = 84/91 = 92\%$

Prevalencia = $(a + c)/(a + b + c + d) = 104/195 = 53\%$

TABLA 4

PRUEBA DE ESFUERZO COMO PREDICTOR DE ESTENOSIS CORONARIA EN PRESENCIA DE UNA PREVALENCIA DEL 17% (1)

| | | > 75% de estenosis | | |
|--------------------|----------|--------------------|---------|-----|
| | | Presente | Ausente | |
| Prueba de esfuerzo | Positiva | 55 | 42 | 97 |
| | | a | b | |
| | Negativa | 49 | 474 | 527 |
| | | 104 | 520 | 624 |

Valor predictivo positivo = $a/(a + b) = 55/97 = 57\%$

Valor predictivo negativo = $d/(c + d) = 478/527 = 91\%$

Sensibilidad = $a/(a + c) = 55/104 = 53\%$ (Como en la tabla 3)

Especificidad = $d/(b + d) = 84/91 = 92\%$ (Como en la tabla 3)

Prevalencia = $(a + c)/(a + b + c + d) = 104/624 = 17\%$

Ahora bien, se dijo que este resultado podía predecirse de la tabla 3, y es esta característica predictiva la que permite extrapolar al lector los resultados de la evaluación de una prueba diagnóstica a su propio medio. Sólo se necesita una estimación gruesa de la prevalencia de la enfermedad en el medio propio (basada en la experiencia personal) o en medios similares (proveniente de otros artículos) y un poco de aritmética simple. Por ejemplo, según lo estimado para la tabla 4, aproximadamente en una sexta parte de todos los sujetos varones (tanto sintomáticos como asintomáticos) referidos de una clínica

de 33% y 88%, respectivamente. Los autores concluyeron: "en mujeres, una prueba de esfuerzo positiva es de limitado valor en la predicción de presencia de enfermedad coronaria significativa, mientras que una prueba negativa es bastante útil en descartar la presencia de enfermedad significativa".

de primer nivel para la práctica de una arteriografía coronaria puede llegar a encontrarse estenosis coronaria. Como resultado, si se parte de los 104 pacientes originalmente considerados, cinco veces este número o 520 pacientes, no tendrían enfermedad. Dado que la sensibilidad permanece constante, 55 (53%) de los 104 varones enfermos presentarían una prueba de esfuerzo positiva. En forma similar, dado que la especificidad permanece siendo del 92%, 478 de los 520 varones sanos tendrían pruebas negativas. Teniendo estos datos, el resto de la tabla puede completarse con sumas o restas simples, calculándose subsecuentemente los valores predictivos

y la exactitud. En esto u otro ejemplo, el valor predictivo positivo se reduce y el valor predictivo negativo aumenta cuando una prueba diagnóstica desarrollada para pacientes con una prevalencia alta de la enfermedad de interés es aplicada en pacientes con una prevalencia más baja de la misma alteración.

La importancia de este análisis deriva de las diferentes prevalencias que tienen las enfermedades en centros de primer y tercer nivel de atención. No obstante, los clínicos rara vez trabajan en más de un nivel de especialización, pudiendo asumirse en tal caso que un clínico dado no necesita preocuparse sobre el efecto que pudieran tener los cambios en prevalencia de una enfermedad en su interpretación de las pruebas diagnósticas. Sin embargo, esta suposición es bastante incorrecta. Ya se han mencionado las diferentes prevalencias entre hombres y mujeres de un mismo medio clínico. Con frecuencia, los pacientes presentan características clínicas fácilmente discernibles que permiten establecer un diagnóstico presuncional en forma bastante precisa antes de la solicitud de cualquier prueba diagnóstica. Por ejemplo, un hombre de 30 años con antecedentes de dolor precordial no anginoso tiene una probabilidad baja de tener estenosis coronaria (Diamond y Forrester (6) estimaron que esta probabilidad es como del 5%); por el contrario, un hombre de 62 años con una angina de pecho típica tiene una probabilidad muy alta de estenosis coronaria (94%) (6). Cuando estas "probabilidades antes de la prueba" o "prevalencias" son incorporadas en el modelo diagnóstico de la prueba de esfuerzo, la información brindada por este estudio varía enormemente. Para un hombre joven puede calcularse que la probabilidad de estenosis coronaria es de 26% si la prueba de esfuerzo es positiva (valor predictivo positivo) y de 3% si la prueba es negativa, (este es el complemento del valor predictivo negativo o $d/[c + d]$). La prueba de esfuerzo es de poco valor en este caso, dado que una prueba negativa sólo informa lo obvio (la enfermedad isquémica del corazón es poco probable en este hombre) y una prueba positiva no implica una probabilidad suficientemente alta de presencia de enfermedad para justificar un estudio más invasivo.

La prueba de esfuerzo tampoco es muy útil en el paciente masculino de 62 años de edad con dolor típicamente anginoso. Si la prueba es positiva, la probabilidad de enfermedad sólo se incrementa del 94% al 99%. Si la prueba es negativa, la probabilidad sólo se reduce al 89%, lo cual difícilmente justifica un mayor escrutinio diagnóstico.

El mejor empleo de la prueba de esfuerzo (o de cualquier otra) está en su aplicación en casos de duda diagnóstica. Considérese otro ejemplo, el de un hombre

de 45 años de edad con angina atípica. Los estudios clínicos demuestran que tal paciente tiene una probabilidad de 46% de tener una estenosis coronaria (6). ¿Debe someterse este paciente a una angiografía? Si se practica una prueba de esfuerzo y ésta es positiva, la probabilidad de enfermedad cardíaca isquémica resulta ser del 05% y justifica la realización de una angiografía. Sin embargo, si una prueba de esfuerzo es negativa, la probabilidad de una estenosis coronaria significativa se reduce a 30%, disminuyendo de esta manera la necesidad de una mayor investigación. En consecuencia, la prueba de esfuerzo es útil, pero sólo en pacientes seleccionados, en quienes la probabilidad de enfermedad coronaria no es ni alta ni baja. Tendría poco sentido actuar de acuerdo a los resultados de la prueba de esfuerzo en el caso de las dos circunstancias previas, ya que en ambas proporcionará poca información adicional.

Habiéndose discutido la comparación con un estándar ideal, ¿qué se puede decir respecto al elemento "a ciegas"? Esto se refiere simplemente a que aquellos que están aplicando o interpretando los resultados de una prueba diagnóstica no deben saber si el paciente evaluado tiene o no tiene la enfermedad de interés; esto es, ellos deben estar "ciegos" al estado de salud real del paciente. En forma similar, aquellos que aplican el estándar ideal no deben conocer los resultados de la prueba diagnóstica en estudio. Sólo cuando la prueba diagnóstica y el estándar ideal son aplicados en forma ciega puede asegurarse que se ha evitado el sesgo consciente o inconsciente (en este caso el sesgo de "sospecha diagnóstica") (7). Recuérdese que este sesgo ya ha sido discutido en un artículo previo sobre desacuerdo clínico (8).

2. ¿Se incluyó un espectro adecuado de pacientes, es decir, de sujetos con enfermedad leve y grave, tratada y no tratada, e individuos con enfermedades diferentes pero cuadro clínico similar?

Una enfermedad en una etapa en que las manifestaciones clínicas son floridas (v. gr.): artritis reumatoide de larga evolución) usualmente representa menor dificultad diagnóstica que cuando se encuentra en etapas tempranas o se presenta en sus formas leves; el valor clínico real de una nueva prueba diagnóstica reside, con frecuencia, en su valor predictivo dentro del grupo de los casos dudosos. Más aún, el valor diagnóstico aparente de algunas pruebas reside más bien en su capacidad para detectar las manifestaciones de la terapia (como serían los depósitos radio opaco en las masas glúteas de los sífilíticos de edad) más que las de enfermedad, y el lector debe estar seguro de que ambas circunstancias no se confundan.

Finalmente, así como un pato no puede ser confundido con un yak aún en ausencia de análisis cromosómicos, la capacidad de una prueba diagnóstica para distinguir enfermedades inconfundibles no apoya su amplio uso. Nuevamente, el valor real de una prueba diagnóstica reside, con frecuencia, en su habilidad para distinguir enfermedades comúnmente confundibles, sobre todo cuando su pronóstico o tratamiento difieren en forma marcada. Es esta propiedad discriminante la que hace que la determinación de T4 sea tan útil en separar a los pacientes tensos, ansiosos, temblorosos y sudorosos entre aquellos con función tiroidea anormal y los que tienen otros padecimientos.

3. ¿Se describió en forma adecuada el medio donde se desarrolló el estudio al igual que el filtro por el que pasaron los pacientes?

En el capítulo previo, se vio cómo la proporción de sujetos hipertensos con lesiones tratables quirúrgicamente variaba casi diez veces dependiendo de si las mismas pruebas diagnósticas se aplicaban en una consulta general o en un centro de especialización. Dado que "el valor predictivo de la prueba cambia con la prevalencia de la enfermedad de interés, el artículo que se lee debe proporcionar información suficiente acerca del lugar donde se desarrolló el estudio y del filtro de selección por el que pasaron los pacientes, a fin de permitir el cálculo del valor predictivo que tendría la misma prueba diagnóstica de ser aplicada en el medio propio.

La selección de los sujetos control, que: no tienen la enfermedad de interés, también debe ser descrita. Pese a que los técnicos de laboratorio y los empleados pueden ser sujetos control apropiados en las etapas tempranas del desarrollo de una nueva prueba diagnóstica (especialmente con el uso decreciente de los estudiantes de medicina como animales de laboratorio), la comparación definitiva con un estándar ideal requiere de una selección cuidadosa tanto de los pacientes con, como de los pacientes sin, la enfermedad de interés. El lector merece cierta seguridad al atribuir las diferencias observadas en los resultados de una prueba diagnóstica a un mecanismo de enfermedad y no a variables tales como edad, sexo, dieta o migración.

4. ¿Se determinó la reproducibilidad de los resultados de la prueba (precisión) y la de su interpretación (variación de observador)?

La validez de una prueba diagnóstica requiere tanto de la ausencia de una desviación de la realidad (esto es, la ausencia de sesgo) como de la presencia de precisión (la misma prueba aplicada al mismo paciente en iguales condiciones debe llegar al mismo resultado).

La descripción de una prueba diagnóstica debe informar a los lectores qué tan reproducibles se espera que sean sus resultados. Esto es importante sobre todo cuando se requiere experiencia en el desarrollo de una prueba (v. gr.: la ultrasonografía tiene, en la actualidad, una enorme variación en la calidad de sus resultados cuando es realizada por operadores diferentes) o en su interpretación (recuérdese de un artículo previo, que la variación del observador es el mayor problema en las pruebas radiológicas, electro-cardiográficas y similares) (9).

5. ¿Se definió en forma adecuada al término "normal"?

Si el artículo usa el término "normal", sus autores deberán especificar su significado. Más aún, el lector deberá quedar convencido de que su definición es clínicamente razonable. En medicina se utilizan varias definiciones de normalidad; algunas de ellas llevan probablemente más a un perjuicio que a un beneficio. Se enlistan seis definiciones de normalidad en la tabla 5, la mayoría de ellas señaladas por Tony Murphy (2,10).

Posiblemente, la definición más común de normalidad asume que los resultados de una prueba diagnóstica (o alguna variante aritmética de los mismos) en un grupo de sujetos supuestamente normales o en una población de "referencia" cuidadosamente caracterizada, se ajustarán a una distribución teórica específica conocida como la distribución normal o gaussiana. Una de las propiedades de la distribución gaussiana es que su promedio \pm dos desviaciones estándar (DE) engloba al 95% de las observaciones, dejando un 2.5% en el extremo superior y otro tanto en el extremo inferior. De aquí que, la expresión "promedio \pm 2DE" se convirtió en una manera tentadora de definir la normalidad y fue así que llegó a hacerse de uso general.

Fue malo que llegara a serlo, ya que tres consecuencias lógicas de su uso han conducido a una enorme confusión y a la creación de una nueva área médica: el diagnóstico de la no-enfermedad (11). Primeramente, los resultados de una prueba diagnóstica simplemente no se ajustan a una distribución gaussiana. (Es más, debería agradecerse que realmente no lo hagan; la distribución gaussiana se extiende al infinito en ambas direcciones, requiriendo ocasionalmente de pacientes con una hemoglobina imposiblemente alta ¡y otros con valores menores a cero!) Segundo, si de todos los valores obtenidos con una prueba diagnóstica se califica como anormal al 2.5% más alto y más bajo, entonces todas las enfermedades tendrían la misma frecuencia, conclusión que clínicamente no tiene sentido alguno.

La tercera consecuencia nociva del uso de la definición, gaussiana de normalidad es compartida con su reciente sustituto: la percentila. Muchos laboratorios, al reconocer el fracaso de los resultados de una prueba diagnóstica en ajustarse a una distribución teórica tal como la gaussiana, sugirieron ignorar la configuración de la distribución y definir como normal, por ejemplo, al 95% inferior de los valores obtenidos. Pese a que la definición de percentila evita el problema de los valores negativos, no evita la conclusión de que todas las enfermedades tienen la misma prevalencia y contribuye al "síndrome del límite superior" de la no-enfermedad, dado que su uso implica calificar como "normales" sólo a aquellos pacientes que no están suficientemente estudiados (2).

Esta consecuencia inevitable resulta de lo siguiente: si el rango de normalidad incluye al 95% inferior de los resultados de una prueba diagnóstica, entonces la probabilidad que tiene un paciente determinado de ser llamado normal después de: someterse a esta prueba es de 0.95 (95%). Si este mismo sujeto se somete a dos pruebas independientes (diferentes en el sentido de estudiar distintos órganos o funciones), la probabilidad de ser calificado como normal sería $0.95 \times 0.95 = 0.90$. De hecho, la probabilidad de que un paciente sea denominado normal es de 0.95 elevado a la potencia del número de pruebas diagnósticas independientes a las que se somete. Por consiguiente, un paciente a quien se le practican 20 pruebas, tiene sólo 0.95^{20} o aproximadamente una posibilidad en 3 de ser llamado normal; un paciente sometido a 100 pruebas tiene aproximadamente 6 posibilidades en 1000 de ser llamado normal, al finalizar su estudio.*

Otras definiciones de normalidad, al tratar de evitar los errores mencionados caen en otros. El punto de vista de "factor de riesgo" se basa en estudios de precursores o predictores estadísticos de eventos clínicos subsecuentes; con esta definición, el rango normal de los niveles de colesterol sérico o de presión arterial estaría dado por los niveles que no implican un riesgo adicional de morbilidad o mortalidad. Desafortunadamente, muchos de estos factores exhiben incrementos constantes en el riesgo asociado a ellos a lo largo de su rango de valores; de hecho, se ha señalado que con esta definición, la concentración normal de colesterol sérico podría encontrarse por debajo de 150 mg/dl (3.9 mmol/l) (13). Otra desventaja de esta definición de factor de riesgo surge al examinar las

consecuencias de una acción ejercida cuando los resultados de la prueba están más allá del rango normal: ¿al alterar el factor se modifica realmente el riesgo? La experiencia reciente en sujetos con niveles "anormales" de colesterol sérico tratados con clofibrato (donde la mortalidad se incrementó —en lugar de descender— con el tratamiento) representa lo peligroso de esta suposición (14).

Un enfoque similar define como normal lo "culturalmente deseable", proporcionando una oportunidad a aquello que Mencken (15) llamó "la corrupción de la medicina por la moralidad" a través de la "confusión de la teoría de los sanos con la teoría de los virtuosos". Desde el punto de vista positivo, tal definición se ve delimitada por la forma de vida actual (v. gr.:) "es mejor ser delgado que obeso" y "el ejercicio y una buena condición física son mejores que la vida sedentaria y una mala condición física" (16); desde el punto de vista negativo, éstas hallan su mejor expresión en el sistema de salud del Tercer Reich. Esta última definición tiene un potencial de daño considerable, y podría llegar al extremo de trastornar el papel de la medicina en la sociedad. A este respecto, Mencken (15) expresa su opinión como: "el verdadero objetivo de la medicina no es el de hacer virtuosos a los hombres; es el de guardarlos y rescatarlos de las consecuencias de sus vicios".

Las dos últimas definiciones son de mucha mayor utilidad para el clínico, dado que se enfocan directamente sobre los actos clínicos de diagnóstico y tratamiento (1). La definición "diagnóstica" delimita un nivel en el resultado de una prueba diagnóstica más allá del cual una enfermedad específica está, con una probabilidad conocida, presente. En ésta la definición que se mencionó en la primera guía de lectura de este capítulo e implica la comparación con un estándar ideal. La "probabilidad conocida" de presencia de una enfermedad es el valor predictivo positivo previamente mencionado.

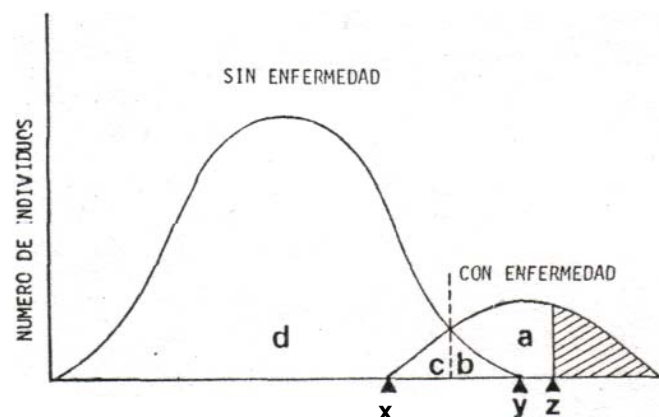
Esta definición se ilustra en la figura 2, donde se observa la sobre posición habitual de los resultados de una prueba diagnóstica de pacientes que, según el estándar ideal, están libres de la enfermedad o enfermos (las letras a, b, c y d de la figura 2 corresponden a las celdillas a, b, c y d de las tablas 2 a 4). La probabilidad conocida (o valor predictivo) con la cual una enfermedad está presente o ausente dependerá de dónde se establezcan los límites del rango normal de la prueba diagnóstica. Si se deseara optimizar el número de veces en que los resultados de una prueba diagnóstica son correctos, se necesitaría ubicar los límites de normalidad con la línea punteada, donde las curvas se cruzan; sin embargo, ello puede no ser muy útil en la clínica.

* Esta consecuencia de las definiciones mencionadas ayuda explicar los resultados obtenidos en un ensayo aleatorio sobre la utilidad del tamizaje con múltiples pruebas realizado al ingreso de los pacientes a un hospital; en dicho estudio no se encontró beneficio para el paciente pero sí un incremento importante en los costos de la atención (12).

TABLA 5

PROPIEDADES Y CONSECUENCIAS DE DIFERENTES DEFINICIONES DE "NORMAL"

| PROPIEDAD | TERMINO | CONSECUENCIAS DE LA APLICACIÓN CLÍNICA |
|--|------------------|--|
| La distribución de los resultados de la prueba diagnóstica tiene una forma característica. | Gaussiano | Ocasionalmente lleva a la obtención de valores negativos de hemoglobina, etc. Todas las enfermedades tienen la misma prevalencia. Los pacientes son normales sólo hasta que son evaluados. |
| Se localiza dentro de una percentila preestablecida obtenida de resultados previos de la prueba diagnóstica. | Percentila | Todas las enfermedades tienen la misma prevalencia. Los pacientes son normales sólo hasta que son evaluados. |
| No implica riesgo adicional en términos de morbilidad o mortalidad. | Factor de riesgo | Infiere que alterando el factor de riesgo se altera el riesgo. |
| Deseable social o políticamente. | Deseable | Confusión sobre el papel de la medicina en la sociedad. |
| Rango de los resultados de una prueba diagnóstica más allá del cual una enfermedad dada está, con una probabilidad conocida, presente o ausente. | Diagnóstico | Necesita del conocimiento de los valores predictivos existentes en determinado medio. |
| Rango de los resultados de una prueba diagnóstica más allá del cual una medida terapéuticamente hace más bien que mal. | Terapéutico | Necesita de una actualización constante de los conocimientos terapéuticos. |



RANGO DE VALORES DE LA PRUEBA DIAGNOSTICA

Fig. 2. Definiciones diagnóstica y terapéutica de "normal"

Si se bajan estos límites normales al punto x, el contenido de la celdilla c alcanza el cero, la sensibilidad y el valor predictivo negativo alcanzan el 100% y el resultado normal de la prueba diagnóstica puede utilizarse para descartar la enfermedad (dado que ninguno de los sujetos enfermos tiene resultados debajo de x). En forma similar, si se suben los límites de normalidad de una prueba diagnóstica al punto y, el contenido de la celdilla b alcanza cero, la especificidad y el valor predictivo positivo alcanzan 100% y el resultado anormal de la prueba diagnóstica puede utilizarse para diagnosticar la enfermedad (dado que ninguno de los sujetos libres de enfermedad tiene resultados superiores a y). De aquí que esta definición tiene utilidad clínica, lo cual la hace superior respecto a las definiciones

previas.

Sin embargo, requiere que el clínico considere tanto los valores predictivos de pruebas diagnósticas individuales como los niveles correspondientes a los puntos x e y aplicables a su propio medio.

La última definición de normalidad ubica sus límites en el punto más allá del cual tratamientos específicos demostraron hacer más bien que mal, y está representada en la figura 2 por el punto z. Esta definición "terapéutica" es atractiva dada su relación con la acción. Por ejemplo, la definición terapéutica del rango normal de presión arterial evita los peligros de etiquetar a los pacientes como enfermos (17), a menos que éstos vayan a ser tratados. El uso de esta definición requiere que el clínico se mantenga actualizado en los progresos terapéuticos y sepa discernir críticamente a los realmente útiles; un artículo subsecuente de esta serie se dedica a este tópico.

En resumen, al leer un informe sobre una nueva prueba diagnóstica, el lector deberá quedar satisfecho con la descripción de lo que los autores definieron como normal, debiendo esta definición haber sido hecha en forma sensata y clínicamente útil.

6. Si la prueba se propone como parte de un grupo o secuencia de pruebas, ¿se determinó su contribución a la validez global?

En muchas condiciones, una prueba diagnóstica individual explora sólo una de varias manifestaciones de una enfermedad. Por ejemplo, en el diagnóstico de la trombosis venosa profunda, la pletismografía por impedancia

examina el vaciamiento venoso, mientras que el rastreo con fibrinógeno marcado con I^{125} explora la actividad de los factores de coagulación en el sitio de la trombosis (10). Más aún, la pletismografía es más sensible en la detección de trombosis venosas proximales que en las distales, mientras que la situación inversa se aplica para el rastreo venoso. Como resultado, la aplicación secuencial de estas pruebas es la más apropiada: si la pletismografía es positiva, se llega al diagnóstico y se inicia tratamiento en forma inmediata; si la pletismografía es negativa, se practica el rastreo venoso y el diagnóstico y las decisiones terapéuticas dependerán del resultado.

Clínicamente no tiene sentido evaluar la utilidad del rastreo venoso comparándolo en forma simple con el estándar ideal de la venografía. La evaluación apropiada de su validez y utilidad clínica debe considerar su concordancia con la venografía en pacientes sintomáticos y con una pletismografía negativa. Otra forma de evaluación apropiada consideraría la concordancia de los resultados combinados del rastreo venoso y la pletismografía con la venografía.

En resumen, cada uno de los componentes de una serie de pruebas diagnósticas debe ser evaluado dentro del contexto de su empleo clínico.

7. ¿Se describieron con suficiente detalle las tácticas para desarrollar la prueba, de manera que ésta pueda replicarse con exactitud?

Si los autores concluyen que su prueba diagnóstica debe ser utilizada, entonces deben decir a sus lectores cómo hacerlo; esta descripción debe considerar tanto las características de los pacientes como los procedimientos de desarrollo de la prueba y de la interpretación de sus resultados. ¿Se requieren condiciones especiales en términos de líquidos, dieta o actividad física? ¿Qué medicamentos deben evitarse? ¿Qué tan doloroso es el procedimiento y qué se hace para aminorar cualquier molestia? ¿Qué precauciones deben ser tomadas durante y después de la prueba? ¿Cómo debe transportarse la muestra y cómo debe almacenarse para análisis subsecuente? Estas estrategias y precauciones deben ser descritas si se desea que los pacientes de uno se beneficien con la prueba diagnóstica propuesta.

8. ¿Se determinó la "utilidad" de la prueba?

El criterio fundamental de una prueba diagnóstica o de cualquier otra maniobra clínica considera los beneficios para el paciente. Si se está de acuerdo con este aspecto, el lector debe analizar el artículo para ver si los autores fueron más allá de lo que a la exactitud, precisión y otras propiedades se refiere, es decir, si se preocuparon de explorar las consecuencias a largo plazo del uso de la prueba diagnóstica.

Además de describir lo que ocurrió con los pacientes clasificados correctamente por la prueba diagnóstica, los autores deben describir lo acontecido a los pacientes que tuvieron resultados falsamente positivos (aquellos con resultados positivos que en realidad no tenían la enfermedad) y aquellos que tuvieron resultados falsamente negativos (los que tuvieron resultados negativos pese a que en realidad tenían la enfermedad). Más aún, si la aplicación de la prueba implica un retraso en el inicio de la terapia definitiva (mientras el procedimiento se reprograma, los tubos con las muestras se incuban o las laminillas esperan ser leídas) las consecuencias de tal retraso deberán ser descritas.

Por ejemplo, considérese el estudio que ha evaluado la utilidad de pruebas no invasivas en el diagnóstico de la trombosis venosa profunda de la pierna, donde se han evaluado también las consecuencias de retardar el inicio de anticoagulantes en aquellos pacientes con una pletismografía por impedancia negativa (una prueba rápida) hasta que, o a menos que, el rastreo venoso con fibrinógeno marcado con I^{125} se hiciera positivo (18). El rastreo venoso tarda horas o incluso días en hacerse positivo cuando los trombos venosos son pequeños o se hallan confinados a la pantorrilla; por ello, es importante determinar e informar si algunos pacientes presentan eventos embólicos durante este intervalo de espera (afortunadamente no los presentan). Más aún, comparaciones de estas investigaciones contra el estándar ideal de la venografía han documentado las consecuencias del tratamiento de pacientes con - resultados falsamente positivos y las consecuencias de su suspensión en pacientes con resultados falsamente negativos. Vale la pena enfatizar el parecido de este enfoque a la definición "terapéutica" de normalidad.*

USO DE LAS GUÍAS EN LA LECTURA

Aplicando las guías expuestas, el lector debe ser capaz de decidir si una prueba diagnóstica es útil en su medio o no lo es, o si simplemente no ha sido evaluada en forma adecuada. Dependiendo del contexto en el cual se desenvuelve el lector, una u otra de las ocho guías será la más importante. Si la guía más importante se satisface en forma confiable, entonces puede pasarse a las otras; si éste no es el caso, el artículo puede ser eliminado sin pasar a aspectos ulteriores. De esta manera, se da mayor eficiencia al uso del tiempo disponible para la lectura. Cuando se busca seleccionar la mejor alternativa diagnóstica de

* A este respecto, es una pena que el término de "eficacia diagnóstica" se haya introducido en la literatura, sobre todo por su empleo como sinónimo de exactitud más que de Utilidad.

un grupo de pruebas competitivas, las guías aquí expuestas facilitan su comparación, y es en base a esta comparación que el lector puede seleccionar la prueba que mejor se ajuste a sus requerimientos.

El siguiente capítulo de esta serie considerará los trabajos sobre curso clínico y pronóstico de una enfermedad.

REFERENCIAS

1. Sackett DL: Clinical diagnosis and the clinical laboratory. *Clin Invest Med*, 1: 37, 1978.
2. Murphy EA: The Logic of Medicine, Johns Hopkins, Baltimore, 117-160, 1976.
3. Ransoloff DF & Feinstein AR: Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med*, 299: 926, 1970.
4. Calen RS & Gombino SR: Beyond Normality: The predictive value and efficiency of medical diagnoses \Wiley, New York, 30-40, 1975.
5. Sketch MU, et al: Significant sex differences in the correlation of electrocardiographic exercise testing and coronary arteriograms. *Am J Cardiol*, 36: 169, 1975.
6. Diamond GA & Forrester JS: Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease. *N Engl J Med*, 300: 1350, 1979.
7. Sackett DL: Bias in analytic research. *J Chronic Dis*, 32: 51, 1979.
8. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont: Clinical disagreement: II. How to avoid it and how to learn from one's mistakes. *Can Med Assoc J*, 123: 613, 1980.
9. ídem: Clinical disagreement: I. How often it occurs and why. *Ibid*: 499-504.
10. Murphy EA: The normal, and the perils of the sylleptic argument. *Perspect Biol Med*, 15: 566, 1972.
11. Meador CK: The art and science of nondisease. *A' Engl J Med*, 272: 92, 1965.
12. Muirbridge TC, et al: An evaluation of multiphasic screening on admission to In-pital. Precis of a report to the National Health and Medical Research Council. *Med J Aust*, 1: 703, 1976.
13. Kannel WB, et al: Preliminary report: The determinants and clinical significance of serum cholesterol. *Mass J Med Technol*, 4: 11, 1962.
14. Oliver MF, et al: A co-operative trial in the primary prevention of ischaemic heart disease using clofibrate. Report from the Committee of Principal Investigators. *Br Heart J*, 40: 1069, 1978.
15. Minchen HL: A Menckon Chrestomathy. Knopf, Westminster, 1949.
16. Lalonde M: A New Perspective on the Health of Canadians. A Working Document Department of National Health and Welfare, Ottawa, Apr 58 1974.
17. Haynes RB, et al: Increased absenteeism from work following the detection and labelling of hypertensives. *N Engl J Med*, 299: 741, 1978.
18. Hull R, et al: Combined use of leg scanning and impedance plethysmography in suspected venous thrombosis. An alternative to venography. *N Engl J Med*, 296: 1497, 1977.

COMO LEER REVISTAS MEDICAS

III. Para conocer el curso clínico y pronóstico de la enfermedad*

DEPARTAMENTO DE EPIDEMIOLOGÍA CLÍNICA Y BIOESTADÍSTICA,
UNIVERSIDAD McMASTER, HAMILTON-ONTARIO, CANADÁ

Esta es la tercera parte de una serie dedicada a desarrollar estrategias y tácticas efectivas, a la vez que eficientes, para leer revistas médicas. Esta parte se refiere al propósito de leer estas revistas para tener conocimiento del curso clínico y del pronóstico de cierta enfermedad. Como en los capítulos anteriores, las reglas de lectura comienzan por los cuatro principios generales de examinar el título, los autores, el resumen y el sitio de realización del estudio (Fig. 1). Después, las reglas son diferentes dependiendo del objetivo del lector.

PRESENTACIÓN DE CASOS

Un ingeniero de 32 años de edad, sano por lo demás, expulsa (y desecha) un cálculo urinario. Ni el interrogatorio ni la exploración física revelan datos significativos y el análisis de orina y la cuantificación de calcio sérico resultan normales. Usted se pregunta si el riesgo de recurrencia es suficientemente alto para justificar una evaluación más profunda.

Una robusta niña de 12 años es enviada a su consulta dado que en un examen rutinario en su escuela se le encontró escoliosis de 10 grados. Una vez que usted corrobora el hallazgo, los padres de la niña le expresan su preocupación de que en el futuro ella llegue a ser una inválida.

Un contador de 37 años tiene colitis ulcerativa crónica, bien controlada, de 20 años de evolución. A pesar de que el proceso se encuentra confinado al colon izquierdo, un colega lo ha enviado a usted para que considere la conveniencia de realizar una colectomía para impedir el desarrollo de

futuro de cáncer colorrectal.

Comentario

Un aspecto central para el manejo clínico de estos pacientes consiste en reflexionar sobre el curso más probable que estas enfermedades seguirán en el tiempo: ¿recurrirá el paso de cálculos urinarios, o ha sido el episodio una manifestación de alguna enfermedad grave que requiere de diagnóstico inmediato? ¿Progresará la escoliosis de 10 grados a un estado de serio deterioro físico o fisiológico? ¿Tienen los pacientes con colitis ulcerativa que afecta el lado izquierdo un riesgo mayor de desarrollar cáncer colorrectal? Estos cuestionamientos son fundamentales al decidir sobre la necesidad de alguna intervención en estos casos y para planear una explicación y consejo apropiados para los enfermos y sus familias.

Cada uno de estos pacientes está experimentando la historia natural de su enfermedad -es decir-, el proceso de interacción del sujeto con los factores causales y el resto del medio ambiente, que ocurre a lo largo del tiempo, y que comienza con el inicio biológico de la enfermedad y termina con la recuperación, la muerte o algún otro estado físico, social o emocional. Para los clínicos resulta de especial interés aquella parte de esta historia natural que comienza con el primer signo o síntoma inconfundible de enfermedad: el curso clínico. Para ser más precisos, lo que se nos pide es formular juicios sobre el pronóstico, o sea sobre la probabilidad de que uno u otro de los desenlaces posibles ocurran en la historia natural y el curso clínico de una enfermedad.

Solicitud de sobretiros- Unidad de Epidemiología Clínica. c/o Dr. Samuel Ponce de León. Instituto Nacional de la Nutrición Salvador Zubirán, Vasco de Quiroga 15, Delegación Tlalpan, 14000 - México, D. F. México.

* Traducción del original publicado en el Canadian Medical Association Journal, vol. 124; pp. 869-872, 1981.

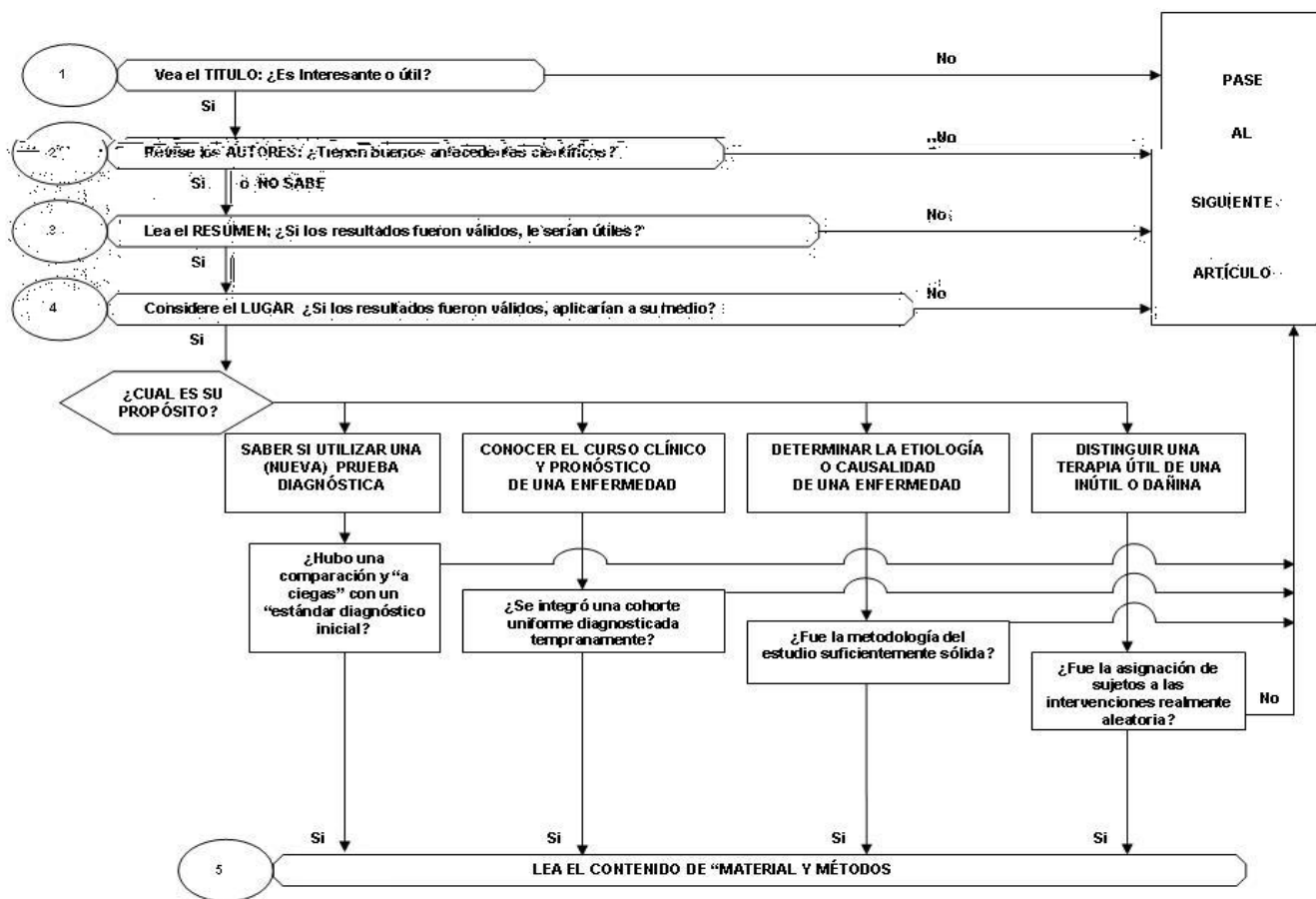


FIG. 1. Los primeros pasos de cómo leer artículos en una revista clínica

ocurran en la historia natural y el curso clínico de una enfermedad.

Para tener información acerca del curso clínico y el pronóstico de la enfermedad podemos buscar artículos-médicos que se refieran a estos puntos o podemos preguntar a alguien que los haya leído, pero ¿qué valor podemos adjudicar a esos escritos? En la literatura, las tasas de recurrencia de cálculos urinarios varían desde 40% (1) (en cuyo caso podríamos abstenemos de un estudio más profundo después de un primer cálculo) hasta casi 100% (2) (situación en la que sería preferible llevar a cabo un estudio a fondo). Las variaciones en las tasas publicadas de recuperación espontánea en casos de escoliosis son del orden de siete veces entre la menor y la mayor (3), y para el riesgo de cáncer en pacientes con colitis ulcerativa son de tres veces, desde 3% (4) hasta casi 10% (5). Ante tanta inconsistencia en los artículos médicos uno se pregunta: ¿cómo es que podemos usarlos sensatamente para hacer decisiones clínicas, acerca del curso clínico y pronóstico de la enfermedad?

Buena parte de las inconsistencias entre los estudios publicados sobre el curso clínico y pronóstico se deben a las diversas formas (muchas de ellas erróneas) en que los autores seleccionaron y siguieron a los pacientes. Sin embargo, los requerimientos para el estudio apropiado del curso clínico y pronóstico son relativamente sencillos, incluyen ese "sentido común aplicado" que se ha mencionado en capítulos anteriores de esta serie, y pueden resumirse en un breve conjunto de normas o guías de lectura que puedan ser aplicadas rápidamente a los artículos publicados (6).

GUIAS PARA LEER ARTÍCULOS SOBRE EL CURSO CLÍNICO Y PRONÓSTICO DE LA ENFERMEDAD

Para mayor facilidad, las seis guías pueden ser enunciadas en forma de preguntas, que se listan en la tabla 1. Puesto que su foco de acción es la forma en que el estudio fue realizado, deben aplicarse a la sección de *Material (o Pacientes) y Métodos*.

TABLA I

GUIAS PARA LEER ARTÍCULOS SOBRE EL CURSO CLÍNICO Y PRONÓSTICO DE LA ENFERMEDAD

1. ¿Se reunió una "cohorte incipiente"?
2. ¿Se describió el patrón de referencia de casos?
3. ¿Se logró un seguimiento completo?
4. ¿Se desarrollaron y usaron criterios objetivos para evaluar el desenlace?
5. ¿La evaluación del resultado o desenlace ¿e hizo en forma ciega?
6. ¿Se tuvo en cuenta el papel de otros factores pronósticos pertinentes?

1. ¿Se reunió una "cohorte incipiente"?

Los pacientes deben ser identificados en un punto uniforme y temprano ("incipiente") del curso de su enfermedad. Este puede ser, por ejemplo, cuando surge la primera manifestación de la enfermedad no atribuible a otro problema o cuando el paciente recibe su primer tratamiento definitivo (con intento curativo). Así se asegura que aquellos que sucumban o se recuperen por completo sean incluidos junto con quienes persistieron con la enfermedad.

Muchos estudios sobre pronóstico se realizan en un sentido inverso. Por ejemplo, varios estudios sobre el riesgo de recurrencia de cálculos urinarios indagan si pacientes sintomáticos en ese momento han arrojado cálculos con anterioridad. Al actuar así, no se presta atención al hecho de que los formadores recurrentes de cálculos (con antecedentes positivos) tienen múltiples oportunidades de ser incluidos en tales series; por el contrario, los enfermos sin recurrencia (con antecedentes negativos) sólo tienen una oportunidad de ser incluidos. ¡No es sorprendente que las tasas de recurrencia difieran considerablemente entre los diversos estudios! Asimismo, en un estudio reciente sobre el riesgo de cáncer en pacientes con colitis ulcerativa, la razón para incluir a varios casos fue precisamente el desarrollo de cáncer en ellos (5).

Es claro que estos enfermos no habían sido seguidos desde el inicio de la colitis y, puesto que los pacientes con colitis que no desarrollaron el cáncer no hubieran entrado al estudio a través de ese procedimiento de selección, el riesgo aparente de cáncer en esta entidad se elevaría artificialmente (6).

Las consecuencias de no reunir una cohorte incipiente para un estudio de curso clínico y pronóstico son impredecibles. En los problemas que se plantearon en la introducción a este capítulo el efecto sería hacer aparecer el pronóstico más pesimista de lo que es mi realidad. Sin embargo, también puede ocurrir el error del pronóstico de pacientes con infarto del miocardio opuesto.

Supóngase que nos interesa saber más acerca y que para ello leemos un artículo acerca de un conjunto de enfermos con este problema que fue reunido en una unidad de cuidados intensivos para coronarios. Un serio defecto de tal estudio es que no incluiría a aquellos sujetos con infarto que mueren antes de poder llegar a la unidad, no obstante que estos pacientes constituyen más de la mitad de las muertes que ocurren dentro del primer año desde el inicio de un infarto al miocardio. Como resultado, y en contraste a los ejemplos previos, la conclusión de este estudio sería la de un pronóstico falsamente optimista.

Por lo tanto, es un error fatal, en estudios de pronóstico, el no reunir una cohorte incipiente apropiada, cíclos pacientes que están en un punto temprano y uniforme del curso de su enfermedad. De esta forma, la aplicación de esta norma a un artículo puede aumentar su eficiencia al leer revistas médicas: Si los autores no reunieron una cohorte incipiente, deseche el artículo de inmediato y pase a otra cosa.

2. ¿Se describió el patrón de referencia de casos?

Deben describirse las vías de ingreso de los pacientes a la muestra en estudio. Esto es necesario para que el lector sea capaz de decidir si los resultados se aplican a los pacientes que acostumbra ver en su práctica clínica. ¿Vienen los casos de un centro de atención primaria (o del primer nivel)? ¿Fueron vigilados todos los hospitales de una cierta región en búsqueda de casos? ¿Se reunió a los pacientes en un centro de atención terciario, que atrae a los desahuciados, a los ricos o a los casos raros?

Los estudios de curso clínico y pronóstico de la enfermedad frecuentemente tropiezan al decidir el mecanismo de reunión de los pacientes. Son cuatro a los principales tipos de sesgo que pueden introducirse en esta etapa: (6, 7) a) *Sesgo centrípeto*: La reputación de un centro médico importante proviene en parte de su pericia en un área especializada de la clínica; por lo mismo, en ese sitio se concentrarán casos problema que se piense pueden beneficiarse de tal experiencia, b) *Sesgo de popularidad*: En este tipo de centros los expertos pueden admitir y vigilar preferentemente a los casos interesantes o que representan un reto en alguna forma, en vez de aquellos casos comunes o poco problemáticos, c) *Sesgo por filtros de referencia*: La selección que ocurre en cualquier etapa del proceso de referencia (es decir, desde la aparición de la primera manifestación que lleva a consultar a un médico, hasta que llega al sitio de la investigación) puede generar muestras cíclos enfermos en los centros de nivel terciario que sean sustancialmente diferentes de aquellos que se encuentran en la población general.

* N. del T.: Desde el punto de vista epidemiológico, el término cohorte se refiere a un conjunto de sujetos (enfermos o no) que son identificados y reunidos en un momento biológico, equiparable a todos, y que a partir de entonces son seguidos a lo largo del tiempo.

d) *Sesgo de accesibilidad diagnóstica:* Los pacientes difieren en su capacidad económica y ubicación geográfica para poder alcanzar la tecnología clínica que permite identificarlos como elegibles para estudios del curso clínico

Lo anterior nos lleva a la siguiente guía:

5. ¿La evaluación del resultado o desenlace se hizo en forma ciega?

La búsqueda de eventos pronósticos importantes debe hacerse por clínicos a quienes se les haya omitido, deliberadamente, información sobre otros rasgos de los pacientes (es decir, que hayan sido cegados a dichos rasgos). En esta forma se pueden evitar dos fuentes adicionales de sesgo (6, 7) : a) *Sesgo por sospecha diagnóstica*: el clínico que sabe que un paciente tiene un factor pronóstico de supuesta importancia puede llevar a cabo exploraciones más frecuentes o detalladas en busca del evento pronóstico relevante, b) *Sesgo por expectación*: El conocimiento previo de las características clínicas de un caso puede influenciar dramáticamente los juicios de patólogos u otros especialistas que interpretan estudios con fines de diagnóstico.

El sesgo por expectación ha sido ejemplificado en la discusión de epidemiología clínica antes referida (8). Aún en la práctica de rutina, a fin de atenuarlo, muchos centros clínicos reconocen la necesidad de una evaluación ciega inicial de pruebas diagnósticas como electrocardiogramas y radiografías (9). El sesgo por sospecha diagnóstica puede evitarse sujetando a *todos* los pacientes en la investigación a las *mismas* pruebas diagnósticas, quizá a intervalos fijados de antemano y, por supuesto, al final del estudio.

¿Existe algún problema cuando el desenlace es la muerte? Evidentemente no se requiere cegar para evaluar un resultado tan "riguroso". Sin embargo, si bien registrar el *suceso* de la muerte es una tarea nada ambigua (y por tanto no amerita realizarse en forma ciega), el asignar una *causa* a la muerte sí está sujeto a los sesgos por expectación y sospecha diagnóstica y, por ende, debe hacerse en forma ciega.

6. ¿Se tuvo en cuenta el papel de otros factores pronósticos pertinentes?

Supongamos que usted quería saber si la duración de la colitis ulcerativa de su paciente era una determinante importante del riesgo de cáncer. Para obtener una respuesta firme, a usted le gustaría estar seguro (a) de que no hubiera interferencia de otros factores que podrían afectar tanto la duración de la enfermedad como su pronóstico (v. gr.: edad más temprana al inicio, más cursos de un amplio espectro de tratamientos, etc.). El no cumplir con esta norma podría resultar en atribuir roles causales a factores que

son sólo "marcadores" de otros con importancia real (y que no fueron identificados apropiadamente).

Generalmente, es imposible que un lector clínico esté suficientemente familiarizado con el zapateado matemático empleado para ajustar los datos de acuerdo al efecto de, otros factores pronósticos, como para decir si los autores deberían haber usado un análisis de función discriminante en vez de una regresión logística múltiple a pasos. Esta es una situación en la que usted debe, (y usualmente puede) confiar en que lo.* editores de la revista hayan hecho que el artículo fuera revisado y aprobado por un buen bioestadístico. Sin embargo, si no se intentó llevar a cabo ningún ajuste por otros factores pronósticos, es poco probable que el artículo haya sido revisado minuciosamente desde el punto de vista metodológico; en este caso, el lector es el responsable de decidir si era necesaria alguna forma de ajuste de los datos (o tomar en cuenta el papel de otros factores pronósticos).

USO DE ESTAS GUIAS PARA LA LECTURA

En este capítulo se han desarrollado seis guías que un clínico atareado puede aplicar a un artículo que verse sobre el curso clínico y pronóstico de la enfermedad. La aplicación de estas guías debe tener dos resultados: a) Usted puede desechar pronto muchos, si no es que todos, de los artículos que encuentre sobre este tema, lo que incrementa la eficiencia del precioso tiempo que dedica a la lectura. Esto ocurrirá especialmente si aplica con rigor la primera guía (¿Se reunió una cohorte incipiente?), puesto que la mayoría de los estudios defectuosos sobre este tópico no satisfacen este requisito inicial, b) Los artículos que cumplan con las guías le proporcionarán información pronostica válida, consisten le y aplicable en su propia práctica clínica.

El siguiente capítulo de esta serie considerará las normas para leer artículos médicos dedicados a la búsqueda de las causas de enfermedad en el hombre.

REFERENCIAS

1. Almy B, Meirik O & Schonebeck J.- Incidence, morbidity and complications of renal and ureteral calculi in a well defined geographical area. *Scand J Urol Nephrol*, 9: 249, 1975.
2. Coe FL, Keck & Norton ER: The natural history of calcium urolithiasis. *JAMA*. 238: 1519, 1977.
3. Rogala EJ, Drummond DS & Gurr J: Scoliosis: incidence and natural history. A prospective epidemiological study *J Bone Joint Surg (Am)*, 60: 173, 1970.
4. Mottet NK; Neoplasia sequelae. En: *Histopathologic Spectrum of Regional Enteritis and Ulcerative Colitis*, Saunders. Philadelphia. 1959: 217-235.
5. Greenstein AJ, Sachar DI, Smitli II, Pucillo A, Paia-testai AE, Krerl I, Geller SA, Janowitz HD & Aujes

- Att Ir*: Cáncer in universal and left-sided ulcerative colitis: factors determining 'risk. *Gastroenterology* 77: 290, 1979.
6. *Sackett DL & Whelan G*: Cáncer risk in ulcerative colitis: scientific requirements for the study of prognosis (E). *Gastroenterology*, 78: 1632, 1980.
 7. *Sackett DL*: Bias in analytic research *JChronic Dis* 32: 51, 1979.
 8. Department of Clinical Epidemiology' and Biostatistics, McMaster University, Hamilton, Ont: Clinical disagreement: I. How often it occurs and why. *Can Med Assoc J*, 123:449, 1980.
 9. *Spodick DH*: On expert and expertise: the effect of variability in observer performance. *Am J Cardiol*.36: 592, 1975.

COMO LEER REVISTAS MEDICAS

IV. Establecer etiología o causalidad*

DEPARTAMENTO DE EPIDEMIOLOGÍA CLÍNICA Y BIOESTADISTICA,
UNIVERSIDAD McMASTER, HAMILTON-ONTARIO, CANADÁ

Este es el cuarto de una serie consecutiva de artículos de Epidemiología Clínica sobre estrategias y tácticas para leer revistas médicas. Las primeras cuatro guías que se muestran en la figura 1 son aplicables a cualquier artículo (considere el título, los autores, el resumen y el sitio). La quinta guía depende de la razón por la cual el artículo es leído, siendo la del presente artículo, la de aprender más acerca de la etiología (o causa) de enfermedad en el humano.

La revisión de una muestra de publicaciones del volumen 121 del *Canadian Medical Association Journal* muestra que sus lectores clínicos se enfrentan frecuentemente a planteamientos acerca de la etiología de ciertas enfermedades. Por ejemplo, se les advierte sobre las siguientes asociaciones: la fibra de la dieta y el cáncer de colon (1), la cimetidina y la disminución en la libido (2), el fumar y el riesgo de cáncer pulmonar ocupacional (3), el consumo abundante de té y la deficiencia de hierro (4), así como sobre drogas que causan dependencia (5), se propone a la encefalitis viral como una causa de corea de Huntington (6), se menciona al mal funcionamiento de la grasa café como causa de obesidad (7), se implica al *Campylobacter* en la etiología de la ileocolitis (8) y a la ceguera a colores en conductores de automóviles como causa de accidentes de tránsito (9); y finalmente, se discute sobre el papel de los "enfriamientos" y de palear nieve, como causas de muerte súbita (10).

Estos 10 "casos sobre causalidad" tienen implicaciones importantes para los médicos, tanto en la práctica pública como en la privada, debido a que conlleva a dar recomendaciones, tales como: cambiar de hábitos dietéticos de los pacientes, reconsiderar la prescripción de

ciertos medicamentos, restringir la ingesta de té en ciertos grupos poblacionales, prohibir conducir vehículos a gente con ceguera a los colores, o bien, impedirles palear nieve a muchos enfermos.

Si a lo anterior agregamos que nuestras pacientes están expuestas a información en la prensa y televisión sobre estas relaciones de causalidad, se entiende que los clínicos se vean frecuentemente obligados a establecer juicios y a dar consejos acerca de ello.

Con el objeto de ayudar a los ocupados clínicos a evaluar un artículo que pregone mostrar causalidad hemos agrupado algunos "principios aplicados de sentido común" que han sido depurados del trabajo de un grupo de especialistas en metodología, principalmente de Austin Bradford Hill (11).

La aplicación de estos principios de sentido común involucra dos pasos. En primer lugar, los lectores deberán escudriñar la sección de Métodos del artículo para evaluar la solidez de los métodos básicos empleados y, en segundo lugar, deberán aplicar al resto del artículo una serie de "pruebas diagnósticas" sobre causalidad.

Primer Paso: Decidir si los métodos básicos usados fueron sólidos o débiles

En ocasiones, usted podrá identificar el método básico utilizado en un estudio a partir del título; en otras, deberá buscarlo en el resumen o en la sección de Métodos. De tal manera que podrá cumplir este primer paso rápidamente sin haber tenido que leer la Introducción o la Discusión.

Este paso se resume en la tabla 1.

Solicitud de sobretiros: Unidad de Epidemiología Clínica, c/o Dr. Samuel Ponce de León, Instituto Nacional de la Nutrición Salvador Zubirán, Vasco de Quiroga 15. Delegación Tlalpan, 14000 - México, D. F. México.

* Traducción del original publicado en el *Canadian Medical Association Journal*, vol. 124; pp. 985-990, 1981.

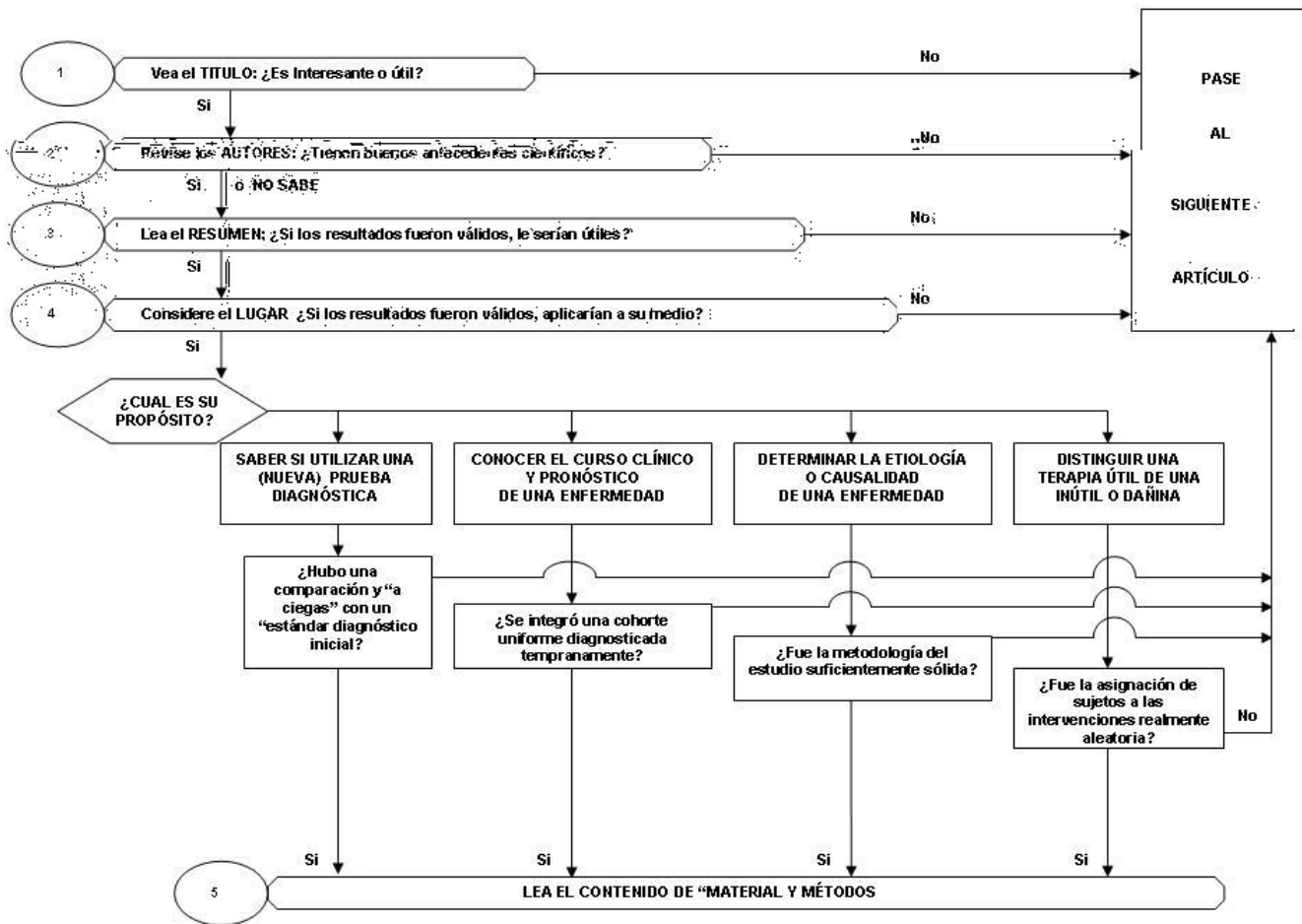


FIG. 1. Los primeros pasos de cómo leer artículos en una revista clínica

Suponga que se desea saber si realmente el palear nieve es causa de ataque cardíaco en hombres de edad media (su edad más 5 años) ; ¿cuál sería el tipo de estudio con mayor solidez que se podría encontrar en la literatura médica?

La mayoría de ustedes, esperamos, comenzaría por buscar un experimento real en humanos: un estudio en el que un grupo de varones en edad media fueran asignados por sorteo (es decir, mediante un sistema semejante al de lanzar una moneda al aire) a palear nieve cada invierno o bien, a no hacerlo,* para después seguirlos y ver cuántos mueren repentinamente en cada uno de los dos grupos. La evidencia sobre causalidad así generada, a través de un *ensayo aleatorio*, es considerada como la más sólida que podamos obtener (ya sea que se trate de aspectos sobre etiología o sobre terapéutica) por razones que si no le son claras aún, le serán conforme avancemos. La arquitectura

* Aquellos que duden de lo factible de este experimento, deben reconocer que el punto que se está tratando aquí es la validez y no la factibilidad..

básica del ensayo aleatorio se muestra en la tabla 2.

Aunque el ensayo aleatorio nos provee con la respuesta más exacta (o válida) a una pregunta sobre causalidad, y es por ende el método con mayor solidez, no lo encontraremos con frecuencia en nuestra lectura clínica. En muchos casos, incluyendo el ejemplo previo, no es factible realizar un ensayo aleatorio para determinar una etiología, e incluso en ocasiones no es ético. Por ejemplo, ¿quién pensaría en llevar a cabo un experimento real en el que en forma deliberada se produjera encefalitis viral a la mitad, asignada por sorteo, de un grupo de individuos con el objeto de ver si tienen un mayor riesgo de desarrollar corea de Huntington? (6).

Por lo tanto, es mucho más probable que encontremos otros estudios llamados sub experimentales, sobre el riesgo de tener un ataque cardíaco por palear nieve. Por ejemplo, después del ensayo aleatorio el método de estudio más sólido que es el *estudio de cohortes*, identificaría a dos grupos (o cohortes) de individuos en edad media: una cohorte que palaria nieve cada invierno y otra que no lo hiciera.

TABLA 1

PRIMER PASO: DECIDIR SI LOS MÉTODOS BÁSICOS UTILIZADOS SON SÓLIDOS O DÉBILES

| SOLIDEZ | MÉTODO |
|---------------|------------------------------|
| El más sólido | Ensayo clínico aleatorio |
| | Estudio de cohortes |
| | Estudio de casos y controles |
| El más débil | Serie de casos |

cada invierno y otra que no lo hiciera. Los investigadores, entonces, seguirían a estas dos cohortes, registrando los ataques cardíacos que ocurrieran en cada una de ellas. En este caso, la dirección del seguimiento es hacia adelante en el tiempo, como se ilustra en la tabla 3. Si la tasa de ataques cardíacos fuera mayor en la cohorte que patea nieve, ésta constituiría una evidencia razonablemente sólida de que al hacerlo se tiene un mayor riesgo de padecer dicho padecimiento. Sin embargo, la solidez de tal estudio de cohortes no es tan grande como la de un ensayo aleatorio. La razón de ello se hace aparente si analizamos al hombre en edad media con angina de pecho. En primer lugar, uno se preguntaría: ¿Será más probable que este hombre evite el patea nieve (u otras actividades que le puedan desencadenar una crisis de angina) que su vecino que no padece de angina do pecho?, la respuesta es: sí.

TABLA 3

ESTRUCTURA BÁSICA DE UN ENSAYO ALEATORIO

| | | Evento (ataque cardíaco) | |
|---|----------------------------------|-----------------------------|----|
| | | SI | No |
| S | Expuestos (palearon nieve) | a | b |
| | No expuestos (palearon nieve) | c | d |

Dirección del seguimiento →

En segundo lugar, nos preguntaríamos si este hombre tiene más riesgo que su vecino de sufrir un ataque cardíaco y nuevamente la respuesta podría ser sí. De esta manera, si por razones ajenas* los individuos con un mayor riesgo de sufrir un ataque cardíaco no estuvieran igualmente

* Usted podrá encontrarse el término *confusor* en sus lecturas y esto es lo que representa la angina de pecho en este ejemplo. En primer lugar es ajena a la pregunta que se hace (¿cuáles son los efectos de patea nieve?); en segundo lugar, es un determinante del evento (ataque cardíaco); y finalmente, se encuentra desigualmente distribuido entre las cohortes de personas expuestas y no expuestas.

TABLA 3

ESTRUCTURA BÁSICA DE UN ESTUDIO DE COHORTES

| | | Evento (ataque cardíaco) | |
|---|---------------|-----------------------------|----|
| | | SI | No |
| ? | Expuestos* | a | b |
| | No expuestos* | c | d |

Dirección del seguimiento →

* Definidos como en la tabla 2.

distribuidos entre las dos cohortes: los que patea nieve y los que no. lo hacen, el estudio analítico de cohortes generaría una respuesta distorsionada a esta pregunta sobre causalidad. Por lo tanto, a un estudio sub experimental, como lo es el estudio de cohortes, deberá tornársele con cierta cautela y desconfianza.

El estudio de casos y controles es un segundo tipo de estudio sub experimental que amerita aún mayor cautela en su interpretación. En este tipo de estudio el investigador selecciona "casos" de hombres que han sufrido un ataque cardíaco y un grupo "control" de individuos que no lo han sufrido. En seguida, a ambos grupos de individuos se les interroga acerca de si patea nieve con regularidad en el invierno. Si la probabilidad de hacerlo fuera mayor en aquellos que sufrieron ataque cardíaco, esto constituiría una evidencia, aunque no muy sólida, de que patea nieve puede causar, o cuando menos precipitar, un ataque cardíaco. Como se muestra en la tabla 4, en este tipo de estudio la dirección del seguimiento es hacia atrás en el tiempo.

TABLA 4

ESTRUCTURA BÁSICA DE UN ESTUDIO DE CASOS Y CONTROLES

| | | Evento (ataque cardíaco) | |
|--|---------------|-----------------------------|-----------|
| | | Casos | Controles |
| | Expuestos* | a | b |
| | No expuestos* | c | d |

Dirección del seguimiento →

* Definidos como en la tabla 2.

¿Cómo es que el estudio analítico de casos y controles ocupa un lugar bajo en la escala de solidez de un estudio?

La razón consiste en que éste es un tipo de estudio muy propenso a estar sesgado. El estudio de casos y controles es susceptible no solamente al sesgo del paciente con angor que mostramos en el estudio de cohortes, sino también a varios otros tipos de sesgos (12). Por ejemplo, si el palear nieve precipitara no solamente ataques cardíacos sino también muerte súbita, sucedería entonces que muchas de estas víctimas no sobrevivirían lo suficiente para ser incluidos en un estudio de casos y controles y mucho menos para ser entrevistados. Como resultado, parecería como si el palear nieve no fuera un factor de riesgo cuando, de hecho, tuvo consecuencias fatales en algunos hombres en edad media. Por ésta, y otras razones similares, los resultados de los estudios de casos y controles son débiles al tratar de aclarar la etiología de enfermedades en el humano.

Por último, haremos mención de otro tipo de estudio sub experimental la *serie de casos*, en la qué el investigador simplemente informa que el 60% de los individuos que tuvieron un ataque cardíaco se encontraban paleando nieve justo antes de sufrirlo. En este tipo de estudio no existe un grupo comparativo y lo único que se puede concluir es que el palear nieve *puede tener* (pero no necesariamente *tiene*) alguna relación causal con los ataques cardíacos. Estas series de casos, aunque pueden generar ideas dan lugar a interpretaciones de más, especialmente por sus investigadores. En términos de solidez la principal utilidad de la serie de casos es el de estimular a realizar otras investigaciones más sólidas, aunque no es raro que dé lugar a que algunos lectores tomen m-83.5o4T5()nleincon-

TABLA 6

IMPORTANCIA INDIVIDUAL DE CADA PRUEBA DIAGNOSTICA EN LA DECISIÓN SOBRE CAUSALIDAD

| Prueba diagnóstica (*) | Efectos del resultado de la prueba en la decisión sobre causalidad (**) | | |
|------------------------------------|---|------------------------------------|--------------------------------|
| | Resultado consistente con causalidad | Resultado neutral o no concluyente | Resultado opuesto a causalidad |
| Experimento en humanos | ++++ | ---- | ----- |
| Solidez de la asociación | | | |
| De un ensayo aleatorio | ++++ | ---- | ----- |
| De un estudio de cohortes | +++ | --- | ---- |
| De un estudio de casos y controles | + | 0 | — |
| Consistencia | +++ | --- | ---- |
| Relación temporal | ++ | --- | ---- |
| Gradiente | ++ | — | --- |
| Sentido epidemiológico | ++ | — | --- |
| Sentido biológico | + | 0 | — |
| Especificidad | + | 0 | — |
| Analogía | + | 0 | 0 |

(*) Enumerados en orden decreciente de importancia.

(**) + := apoya causalidad; — = descarta causalidad; 0 = no afecta la decisión sobre causalidad.

Los signos + y — indican la contribución relativa de la prueba diagnóstica a la decisión sobre causalidad.

especificidad, es decir, es posible "explicar" casi cualquier conjunto de observaciones.

1. ¿La evidencia proviene de un experimento real en seres humanos?

Como explicamos anteriormente, éste es un tipo de investigación en el que grupos idénticos de individuos, generados a través de una asignación por sorteo, estarán o no expuestos al factor causal imputado y serán seguidos para registrar la presentación del evento de interés.

Como acabamos de ver, esta es la mejor evidencia que jamás podremos tener aunque no siempre dispondremos de ella y sólo raramente será una evidencia inicial de causalidad. No obstante, cualquier consideración sobre un aspecto de etiología deberán comenzar por la búsqueda de un ensayo aleatorio.

2. ¿Es la asociación sólida?

La solidez se refiere a la magnitud de la diferencia en la frecuencia del evento de interés entre los expuestos y los no expuestos a la causa imputada. Mientras mayor sea esta diferencia, más fuerte será la asociación.

Existen varias estrategias para estimar qué tan sólida es una asociación. En el ensayo aleatorio y en el estudio de cohortes (tablas 2 y 3) los pacientes que están o no expuestos a la causa imputada son cuidadosamente seguidos para encontrar si la reacción adversa o el evento ocurren. En tal estudio de cohortes se compararía, por ejemplo, la frecuencia de aparición de impotencia entre los pacientes ulcerosos que recibieron cimetidina con respecto a los que no la recibieron (2).

Los estudios de cohortes (tabla 3) son atractivos desde el punto de vista metodológico ya que, al igual que los ensayos aleatorios, permiten el cálculo directo

de la magnitud (riesgo relativo) de la asociación mediante la Comparación de la frecuencia de eventos en las personas expuestas con la de las no expuestas, como sigue:

$$\frac{a}{a+b} \bigg/ \frac{c}{c+d}$$

Sin embargo, como vimos en la sección previa, los estudios de cohortes son frecuentemente prolongados y costosos. Por lo tanto, la mayor rapidez y el menor costo de un estudio de casos y controles (tabla 4), en el que los pacientes con y sin el evento de interés (v. gr. : impotencia) son seleccionados y seguidos hacia atrás en el tiempo para ver su exposición a la causa imputada (v. gr. : cimetidina), lo hacen un enfoque mucho más popular, particularmente como un primer paso para sondear las conclusiones de una serie de casos inicial. Los estudios de casos y controles, o estudios "etrohoc" (16), pagan un precio metodológico a su economía en tiempo y dinero, puesto que la magnitud o riesgo relativo sólo puede ser estimado indirectamente, con la fórmula ad/bc . Este cálculo, aunque algebraicamente justificado, ha sido tomado con cierto escepticismo (16).

Más aún, como hemos visto, los estudios de casos y controles son particularmente vulnerables a una serie de distorsiones sistemáticas (sesgos) que pueden dar lugar a estimaciones erróneas acerca de la magnitud de la asociación y, por lo tanto, a conclusiones /incorrectas en relación a causalidad. Algunos de estos sesgos fueron discutidos en el artículo previo de esta serie (parte III), y aún otros son descritos en detalle en otros artículos, para los lectores que deseen profundizar en ésto (12).

Al revisar los efectos potenciales de estos sesgos en la distorsión de las conclusiones de los estudios de casos y controles y de cohortes, se llega a dos conclusiones. En primer lugar los estudios de casos y controles están sujetos a más fuentes de sesgo que los estudios de cohortes. En segundo lugar, mientras uno puede generalmente anticipar y superar (mediante métodos apropiados y rigurosamente aplicados) los sesgos que afectan a los estudios de cohortes, esta solución es mucho más difícil (o imposible) en la estrategia de casos y controles.

En conclusión, los lectores pueden tenerle bastante confianza a las estimaciones de la magnitud de una asociación proveniente de un ensayo aleatorio, moderada si son de un estudio de cohortes y poca, si son de un estudio de casos y controles.

3. ¿Es la asociación consistente de estudio a estudio?

La demostración repetitiva por diferentes investigadores de una asociación entre la exposición a la causa imputada y el evento de interés, usando estrategias diferentes y en diferentes lugares, es lo que constituye la consistencia. Así, mucho de la credibilidad acerca de la relación causal entre el tabaquismo y el cáncer de pulmón surge de la demostración repetida de una asociación estadística sólida en estudios de casos y controles, de cohortes y en otros diseños.

4. ¿Es la relación temporal correcta?

Para que esta prueba de temporalidad sea positiva se requiere una secuencia consistente de eventos; la exposición a la causa imputada seguida de la aparición del evento de interés. Aunque esta prueba diagnóstica parezca fácil de aplicar, no lo es. ¿Qué pasaría si un segundo factor predisponente, o una etapa muy temprana de la enfermedad, es responsable tanto en la exposición al factor causal imputado como del evento en toda su expresión? De hecho, tal explicación puede aplicarse a estudios que han asociado al uso ilícito de estimulantes o de drogas depresoras con el diagnóstico subsecuente de psicosis o de depresión, respectivamente (17). ¿Fue el uso ilícito de diferentes drogas lo que causó las formas específicas de enfermedad mental diagnosticada posteriormente?, ¿o fueron más bien los individuos, con diferentes estados subclínicos pero progresivos de la enfermedad mental, los que buscaron las drogas específicas? Evidentemente, este criterio es más fácil de aplicar a un estudio de cohortes que a uno de casos y controles, puesto que este último establece una asociación temporal entre la "exposición" y el "evento" después de que ambos ya han ocurrido.

5. ¿Existe un gradiente dosis-respuesta?

Lo que satisface esta prueba diagnóstica es la demostración de un incremento en el riesgo o gravedad del evento de interés, en asociación con un aumento en la "dosis" o duración de la exposición a la causa imputada. Por ejemplo, en un reporte en el que se asocia a los estrógenos conjugados con carcinoma endometrial (18), el riesgo relativo de padecer este se incrementó de 5.6%, en las mujeres que habían usado el fármaco durante 1 a 2.9 años, a 7.2% en aquellas que lo usaron durante 5 a 6.9 años y finalmente, a 13.9% en los que lo habían hecho por 7 años o más.

Los gradientes inversos son útiles también. De hecho, algunas de las evidencias más definitivas acerca de la asociación entre el consumo de cigarrillos y el cáncer pulmonar es la disminución progresiva en el riesgo de cáncer que se ha informado en exfumadores conforme celebran un nuevo aniversario de su último cigarrillo.

6. ¿Tiene la asociación sentido, desde el punto de vista epidemiológico?

Esta guía se cumple cuando los resultados del artículo concuerdan con nuestro entendimiento actual de la distribución de causas y eventos en los humanos.

Por ejemplo, Freeman (1), al revisar el posible papel de la fibra en la dieta en la patogenia del cáncer de colon, registró en varios estudios que la distribución de fibra dietética en diferentes áreas geográficas o poblaciones estaba inversamente relacionada a la aparición de cáncer de colon en las mismas áreas y grupos humanos. Reconociendo la debilidad de tales correlaciones epidemiológicas (después de todo, la disminución en la tasa de nacimiento en Europa se ha acompañado paralelamente de la desaparición de cigüeñas en sus ciudades), Freeman hizo necesarios los "estudios prospectivos a largo plazo" para mejor definir el papel de la fibra en la dieta en el cáncer del humano.

7. ¿Tiene la asociación sentido, desde el punto de vista biológico?

¿Existe concordancia entre la asociación y el entendimiento actual de las respuestas a estímulos, de células, tejidos, órganos y organismos? Es con este patrón como deben ser evaluados los datos experimentales no obtenidos de humanos. Aunque virtualmente cualquier conjunto de observaciones pueden plantearse como biológicamente plausible (dada la ingenuidad de la mente humana y el vasto surtido de hechos biológicos contradictorios) algunas observaciones biológicas pueden ser importantes, tales como la descripción por Himms-Hagen sobre la inducción de obesidad masiva en ciertas capas de ratones cuya grasa café tenía una capacidad limitada de termogénesis (7).

8. ¿Es la asociación específica?

Lo que satisface a esta prueba diagnóstica es la existencia de una asociación entre una sola causa imputada y un efecto único. Como ejemplos están algunas alteraciones genéticas muy características en las que alteraciones en una sola enzima (o alguna otra proteína producen enfermedades específicas tales como la hemofilia A o la cistinuria). Esta es una prueba diagnóstica menor, siendo útil sólo en forma moderada y únicamente cuando la enfermedad está presente. La debilidad de esta prueba es puesta en evidencia si consideramos que los teratógenos comúnmente tienen efectos múltiples y en diversos órganos

9. ¿Es la asociación de causalidad análoga a otra, previamente demostrada?

En esta última y menos importante prueba diagnóstica se liaría la analogía entre el cáncer de escroto en los desolladores de chimeneas, en épocas anteriores, con la aparición, más reciente, del cáncer pulmonar en los individuos que inhalan (en lugar de portar en sus ropas) los productos de la combustión.

Uso de estas guías al leer

Cada vez que usted se enfrente a una pregunta sobre causalidad o etiología puede usar estas nueve pruebas diagnósticas para así refinar su lectura clínica y, con la ayuda de los pesos específicos que se muestran en la tabla 6, alcanzar una conclusión sobre causalidad. Aún antes de leer, usted puede usar estas guías para aumentar la eficiencia de búsqueda bibliográfica ya que se enfocaría a las publicaciones que le dieran las evidencias más sólidas con respecto a su pregunta sobre causalidad y le prevendrían de aceptar conclusiones posibles pero sesgadas.

Sin embargo, aún después de leer exhaustivamente y de aplicar todas las nueve pruebas diagnósticas usted puede aún permanecer incierto acerca de si, por ejemplo, el fármaco A realmente produce la enfermedad B. ¿Qué es lo que haría, entonces, y cómo traduciría esta deliberación en una acción clínica?

Nosotros sugerimos que esta "decisión para actuar" tenga dos componentes (Fig. 2). Primero, es la *certeza sobre causalidad*, que se basa en el resultado de aplicar las nueve pruebas diagnósticas antes mencionadas. Lo segundo, es *considerar las consecuencias de acciones alternativas* que se nos presenten (lo que incluye la *n*-interferencia así como el mantenimiento del *status quo*).

| DECISIÓN CLÍNICA PARA ACTUAR | CERTEZA SOBRE CAUSALIDAD | CONSIDERACIONES SOBRE LAS CONSECUENCIAS DE TODAS LAS ACCIONES ALTERNATIVAS |
|------------------------------|--------------------------|--|
| | = | + |

Así la decisión de actuar o no resultaría de la interacción de estos dos componentes. Considérense los siguientes dos ejemplos.

En 1974 aparecieron en forma abrupta tres informes implicando a la reserpina como una causa de cáncer de mama (19-21), lo que precipitó una crisis en el manejo de la hipertensión. ¿Qué le dinamos y cómo trataríamos a los pacientes cuya presión arterial se controlara con este fármaco? El primer componente de esta decisión consideró el grado de certeza de que la reserpina, en efecto, producía cáncer de mama; este nunca fue sólido (de hecho, esta droga fue luego perdonada por sus acusadores anteriores). Por otra parte, el segundo componente de la decisión identificó una acción alternativa muy atractiva para muchos clínicos canadienses: cambiar la reserpina por propranolol en los pacientes apropiados. Así, en este caso, aún un bajo grado de certeza de causalidad fue tomado en cuenta para la decisión clínica de suspender la reserpina en muchos pacientes, puesto que había una alternativa terapéutica disponible.

En contraste con esto, a pesar de que el grado de certeza cfe que los anticonceptivos orales causan tromboembolismo es mucho más alto, estos fármacos siguen siendo ampliamente usados. El razonamiento que respalda la decisión de continuar su uso ante la evidencia creciente de que causan tromboembolismo es complejo y se debe, al menos en parte¹, al segundo componente de la decisión: se considera que las consecuencias de un control natal alternativo son aún menos deseables que el riesgo pequeño, aunque real, de tromboembolismo. Así el uso de contraceptivos orales continúa, e interesantemente la prueba diagnóstica del gradiente dosis-respuesta está involucrada en la justificación de la reducción progresiva cfe ciertos constituyentes hormonales en los anticonceptivos orales.

El diagnóstico de causalidad no es simplemente aritmético y las estrategias y tácticas para evaluarla son aún primitivas. Las pruebas diagnósticas aquí presentadas son un principio y sugerimos que su uso, particularmente cuando se especifique: con claridad antes de la revisión de datos relevantes, conduzca a discusiones más racionales (aunque menos coloridas) sobre causalidad en medicina.

El siguiente y último artículo en esta serie estará dirigido a cómo leer revistas clínicas que traten sobre la distinción entre una terapia útil y una inútil o incluso perjudicial.

REFERENCIAS

- 1 Freeman HJ: Dietarv fibre and colonic neoplasia. *Can Med Assoc J.* 1979; 121: 291, 2. Biron P: Diminished libido with cimetidine therapy (C). *Ibid:* 404-405.

Fig. 2 Componentes de una "decisión clínica para actuar".

3. *Chovil AC*: Occupational lung cancer and smoking: a review in the light of current theories of carcinogenesis. *Ibid*: 548-555.
4. *Farkas CS*: Body iron status associated with tea consumption (C). *Ibid*: 706.
5. *Jacob MS & Sellers EM*: Use of drugs with dependence liability. *Ibid*: 717-724.
6. *Averback P*: Viral encephalitic pathogenesis of Huntington's chorea? (C). *Ibid*: 1060-1062.
7. *Himms-Hagen J*: Obesity may be due to a malfunctioning of brown fat. *Ibid*: 1361-1364.
8. *Lambert JR, Fischler ME, Karmali MA & Newman A*: *Campylobacter* ileocolitis and inflammatory bowel disease. *Ibid*: 1377-1379.
9. *Ilés JDH*: Colour-blind drivers of motor vehicles (C). *Ibid*: 1566.
10. *Anderson TW & Rochard C*: Cold snaps, snowfall and sudden death from ischemic heart disease. *Ibid*: 1580-1583.
11. *Hill AB*: Principles of Medical Statistics, 9th, ed, Lancet, London, Engl 312, 1971.
12. *Sackett DL*: Bias in analytic research. *J Chronic Dis*, 32: 51, 1979.
13. Veterans Administration Cooperative Study Group on Antihypertensive Agents; Effects of treatment on morbidity in hypertension: III. Influence of age, Diastolic pressure, and prior cardiovascular disease; further analysis of side effects. *Circulation*, 45: 991, 1972.
14. Oral Contraceptives and Health. An interim report from the oral contraceptive study of the Royal College of General Practitioners. Pitman Med, London, Engl, 1974.
15. *Labarthe DR*: Methodologic variation in case-control studies on reserpine and breast cancer. *J Chronic Dis*, 32: 95, 1979.
16. *Feinstein AR*: Clinical bioesthetics, XX. The epidemiologic trohoc, the ablative risk ratio, and "retrospective" research. *Clin Pharmacol Ther*, 14: 291, 1973.
17. *McLellan AT, Woody GE & O'Brien CP*: Development of psychiatric illness in drug abusers. Possible role of drug preference. *N Engl J Med*, 301: 1310, 1979.
18. *Ziel HK & Finkle WD*: Increased risk of endometrial carcinoma among users of conjugated estrogens. *N Engl J Med*, 293: 1167, 1975.
19. Boston Collaborative Drug Surveillance Program: Reserpine and breast cancer. *Lancet*, 2: 669, 1974.
20. *Armstrong B, Stevens N & Doll R*: Retrospective study of the association between use of rauwolfia derivatives and breast cancer in english women. *Ibid*: 672-675.
21. *Heinonen OP & Shapiro S*: Tuominen L Turunen MI: Reserpine use in relation to breast cancer. *Ibid*: 675-677.
22. *Shapiro S & Slone D*: The case control study: consensus and controversy. Comment. *J Chronic Dis*, 32: 105, 1979.

COMO LEER REVISTAS MEDICAS

V. Para distinguir entre tratamientos útiles e inútiles o incluso perjudiciales*

DEPARTAMENTO DE EPIDEMIOLOGÍA CLÍNICA Y BIOESTADISTICA,
UNIVERSIDAD McMASTER, HAMILTON-ONTARIO, CANADÁ

Esta serie, de artículos consecutivo sobre Epidemiología Clínica presenta estrategias efectivas y eficientes para el análisis de la información que los médicos revisan al leer artículos clínicos. Las reglas generales para leer artículos clínicos se resumen en la figura 1. En este último artículo de la serie se revisará la lectura de revistas clínicas para distinguir entre tratamientos útiles y tratamientos inútiles o inclusive perjudiciales; las guías para hacer esto se enlistan en la tabla 1. La necesidad para efectuar esta distinción se enfatiza en las siguientes situaciones clínicas:

A.—Un ejecutivo de 48 años de edad se sometió a una revisión clínica de rutina encontrándose con niveles elevados de colesterol. Su historia clínica es irrelevante y el ECG de reposo es normal. Se le prescriben una dieta baja en grasa y clofibrato.

B.—A un hombre con antecedentes de úlcera duodenal en su padre a quien se trató por congelación gástrica quince años antes también se le encuentra una úlcera duodenal y se le inicia tratamiento con cimetidina.

C.—A una mujer embarazada con antecedentes de abortos espontáneos se le realiza un circlaje.

D.—A un hombre con angina de pecho grave y estenosis marcada de la arteria coronaria izquierda se le realiza un puente aorto-coronario en un servicio quirúrgico en donde previamente se realizaba, ligadura de la arteria mamaria interna.

E.—Después de estar hospitalizada durante 6 meses por esquizofrenia, a una mujer se le da de alta 1:011 una receta por imipramina.

F.—A un niño cu quien; se diagnostica meningitis tuberculosa se le inicia tratamiento inmediato con estreptomycin, isoniacida y rifampicina.

G.—Dos ancianos con ataques de isquemia cerebral transitoria ingresan al hospital. A uno se le realiza una endarterectomia carótidea y al otro se le da tratamiento a largo plazo con ácido acetilsalicílico.

H.—Al encontrar una presión arterial de 140/90 mmHg en una mujer embarazada con cifras tensionales de 110/70, se le inicia tratamiento con metildopa.

I.—A una mujer con poliartritis y factor reumatoide positivo se le indica indometacina.

J.—Un niño en quien se encontró en un examen pre-escolar una pérdida auditiva de 35 decibels es programado para timpanoplastía e inserción tubaria.

En todas estas situaciones los pacientes llenen varios datos comunes: El que los une para la presente discusión es que a todos ellos se les ha dado un tratamiento que intenta prevenir (clofibrato en el caso A), curar (estreptomycin en el caso P), o mitigar (indometacina en el caso I) la enfermedad o padecimiento. Todas las intervenciones descritas se basaron en los resultados de investigaciones básicas en biología* y en comportamiento humano y en series de casos se han documentado claramente evoluciones clínicas excelentes en pacientes que recibieron esos tratamientos.

* Sí aún para la ligadura de la arteria mamaria interna existen estudios que muestran que la inyección de sustancias en el muñón de la arteria mamaria interna pueden recuperarse poco tiempo después en la circulación coronaria (1).

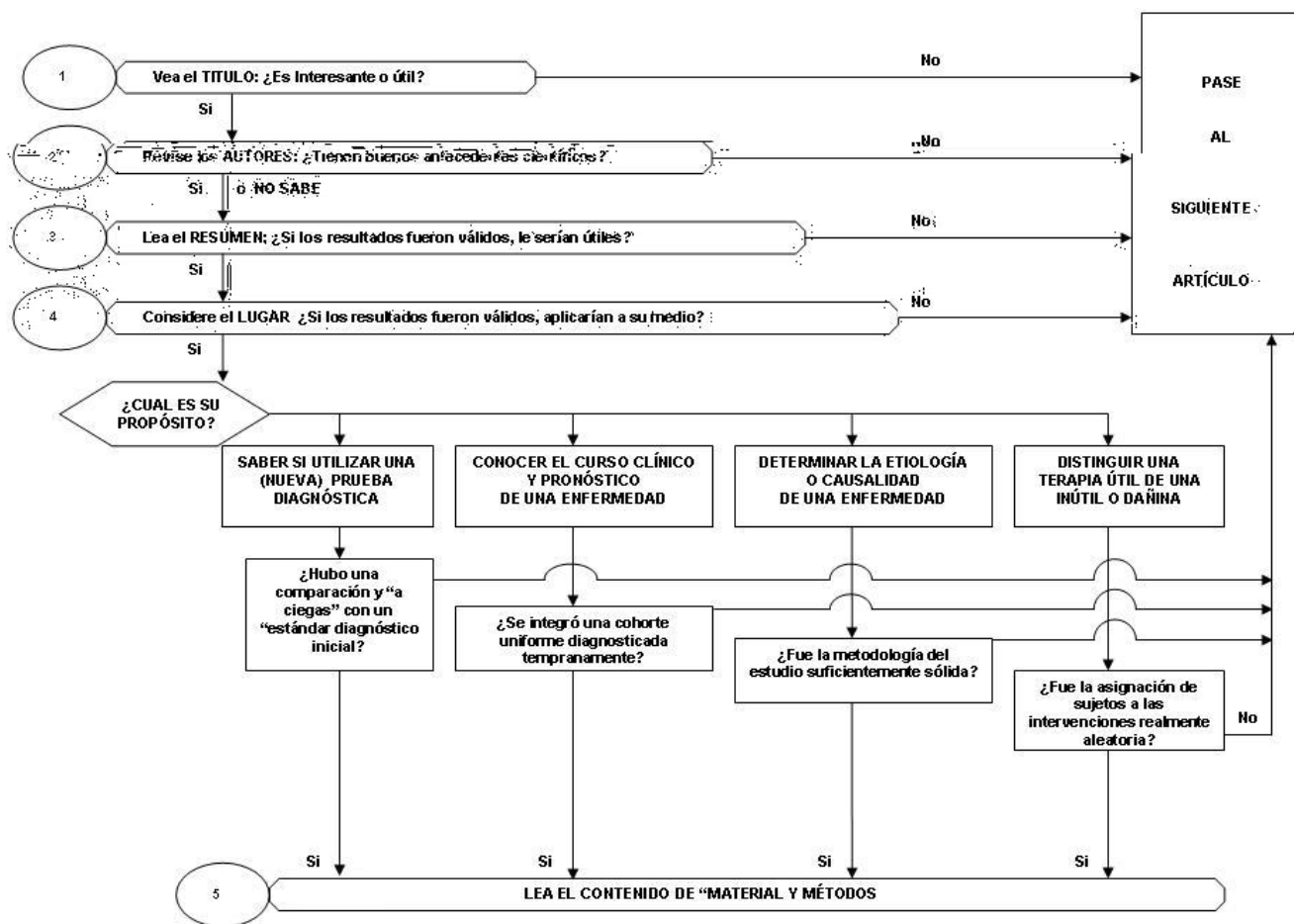


FIG. 1. Los primeros pasos de cómo leer artículos en una revista clínica

Sin embargo, aunque algunos de los tratamientos de los ejemplos son eficaces (esto es, producen más beneficio que daño a los pacientes que los reciben), tenemos serias dudas sobre otros tratamientos y aún más nos burlamos del consenso existente en una era previa que aceptó el congelamiento gástrico y la ligadura de la arteria mamaria interna como procedimientos terapéuticos. ¿Por qué es esto? Una razón principal es que estamos dispuestos a aprender de la experiencia; de esta manera aquellos tratamientos que producen más perjuicio que beneficio son eventualmente desentramados.

Más importante aún es la creciente actitud que reclama que la eficacia de los procedimientos terapéuticos sea respaldada por evidencias sólidas, esto es, con estudios clínicos aleatorios, antes de que los clínicos los consideren satisfactorios y los acepten como modalidades de tratamiento.

En esta discusión se mostrará cómo aplicar algunas reglas de sentido común para evaluar la supuesta eficacia que es preconizada en los artículos médicos que aparecen en las revistas.

GUIAS DEL LECTOR

Las reglas de evidencia científica para, los estudios de terapéutica pueden resumirse en seis puntos fundamentales para el ocupado clínico que lee estos trabajos (tabla 1). Repetimos que éstos constituyen "sentido común aplicado" y se han diseñado para optimizar la eficiencia y la precisión de la lectura. Estas guías son de dos tipos como se observa en la tabla 1: La primera y última se relacionan con *validez* (¿Qué tan probable es que los resultados del estudio sean verdaderos?), y la segunda tercera y quinta guías se relacionan principalmente con la *aplicabilidad* (¿tiene utilidad o resultado del estudio?). La cuarta guía tiene que ver con ambas situaciones tanto validez como aplicabilidad.

1. ¿Fue la distribución de los pacientes en los tratamientos realmente al azar?

Cualquier paciente que entra a un estudio de intervención debería haber tenido la misma probabilidad conocida (característicamente del 50%) de recibir uno de los tratamientos que serán comparados; por lo tanto, la

TABLA.1 UNA VISION DETALLADA DE LAS GUIAS PARA ESTUDIOS DE INTERVENCIÓN

| | | |
|--|--|---|
| VALIDEZ (Qué tan probable es que los resultados sean verdaderos) | 1. ¿Fue la distribución de los pacientes en los tratamientos realmente por azar? | APLICABILIDAD (Qué tan útiles son los resultados) |
| | 2. ¿Fueron informados todos los sucesos clínicos relevantes? | |
| | 3. ¿Fueron los pacientes estudiados similares a los que usted atiende? | |
| | 4. ¿Se consideraron tanto la significancia clínica como la estadística? | |
| | 5. ¿Es el tratamiento propuesto utilizable en su práctica clínica? | |
| | 6. ¿Todos los pacientes que entraron al estudio fueron evaluados al finalizarlo? | |

asignación a uno y otro tratamiento debería ser efectuada por un sistema similar al de "echar un volado" (aventar una moneda al aire). Usualmente es fácil determinar si el estudio se realizó de esta manera dado que en el resumen en la sección de Material y Métodos o incluso en el título aparecerán términos claves como son: "estudio aleatorio" o "distribución aleatoria".

Como resultado de lo anterior el clínico atareado tiene la opción de aplicar rigurosamente esta regla: si se lee para "mantenerse al día" en la investigación clínica relevante, más que escudriñar la literatura en búsqueda de la manera de tratar a un paciente específico, se deberán descartar todos los artículos sobre tratamiento que no sean estudios aleatorios. ¿Por qué usar un criterio tan estricto?, ¿por qué no puede el clínico aceptar los resultados de estudios no aleatorios? Una explicación formal para esta guía sería demasiado extensa pero la conclusión es muy clara: Una distribución aleatoria elimina muchos de los sesgos que llevan a resultados falsos en estudios no aleatorios. La explicación programática es breve; es mucho más probable que beneficiemos a un paciente y mucho menos probable que lo perjudiquemos si utilizamos tratamientos que hayan mostrado ser más benéficos que perjudiciales en ensayos clínicos aleatorios adecuados.

Existen numerosas situaciones en las cuales se han cometido errores por aceptar la evidencia obtenida de estudios no aleatorios y algunas de ellas se incluyen en los casos presentados inicialmente en este artículo.

Por ejemplo, el clofibrato, como se usó en el caso A, tuvo una gran popularidad hasta que un ensayo clínico aleatorio que demostró que este medicamento en efecto aumentaba la mortalidad (2), subsecuentemente esta droga fue prohibida en varios países. Todavía más, se ha estimado que 2,500 máquinas de congelamiento gástrico habían sido utilizadas para tratar decenas de miles de pacientes con úlcera péptica, -por ejemplo, al padre en el caso B- antes de que un estudio aleatorio demostrara la falta de eficacia de este tratamiento (3).

Finalmente, fue necesario un estudio clínico aleatorio en el cual pacientes con angina de pecho fueran aleatoriamente asignados para someterse o no a ligadura de la mamaria interna después de que sus arterias habían sido expuestas quirúrgicamente para demostrar la alta frecuencia de mejoría sintomática que pueden ocasionar los procedimientos o medicamentos placebo (4).

Bywaters señalaba (5), en relación al tratamiento de la fiebre reumática, que los grupos que apoyaban los diversos tratamientos podían dividirse en los que tenían entusiasmo pero no tenían grupo control y los que tenían grupo control pero no tenían entusiasmo. Esta situación fue

TABLA 2

RELACIÓN ENTRE UN SUPUESTO BENEFICIO TERAPÉUTICO Y EL USO DE GRUPOS CONTROL*

| <i>Tipo de estudio</i> | <i>Número de estudios (% de informes de beneficio terapéutico)</i> | |
|------------------------|--|------|
| <i>Sin controles</i> | 19 | (89) |
| <i>Con controles</i> | 18 | (50) |

P = 0.01, Prueba exacta de Fisher

* Modificado de la referencia 6.

En resumen, aunque los ensayos clínicos aleatorios pueden producir conclusiones incorrectas sobre eficacia (especialmente, como veremos en breve, cuando se trata de estudios con números pequeños) es con mucho la mejor herramienta actualmente disponible para identificar las intervenciones terapéuticas que son más benéficas que perjudiciales.

¿En qué situaciones puede uno confiar en que el tratamiento es eficaz sin un estudio aleatorio? Únicamente cuando de manera invariable, los tratamientos convencionales prevengan la muerte de los pacientes. Un ejemplo es el caso F: Antes de 1946, el pronóstico de la meningitis tuberculosa era invariablemente mortal. Cuando se pudo disponer de estreptomycin en los E. U. A., algunas víctimas tratadas con esta nueva droga sobrevivieron (7). Estos resultados fueron confirmados poco después en Inglaterra (8). En estas circunstancias la sobrevivencia de los pacientes tratados con un padecimiento invariablemente mortal constituye, por sí sola, suficiente evidencia de eficacia.

Al poner énfasis en la evidencia obtenida de ensayos clínicos aleatorios usted puede incrementar la eficiencia con la que lee una revista a la que está suscrito, ya que conducirá a rechazar la mayor parte de los artículos. Sin embargo estas reglas requieren de modificaciones cuando uno revisa la literatura buscando información sobre el tratamiento de un paciente particular; frecuentemente en este caso no se encontraran ensayos aleatorios adecuados. ¿Qué es lo que el lector debe hacer ante esta situación?

Cuando se lee sobre un paciente específico hay dos tipos de acciones que son aconsejables. En primer lugar, una revisión inicial en búsqueda de la existencia de algún estudio clínico aleatorio. En segundo lugar y en ausencia de estudios clínicos aleatorios, se deberán considerar los resultados de estudios experimentales. Antes de aceptar las conclusiones de este tipo de estudios, el clínico deberá estar convencido de que la mejoría en la evolución de los pacientes es lo suficientemente importante como para no ser consecuencia del o de los sesgos que pudieran estar presentes en la inclusión de pacientes en el estudio, o en la interpretación del estudio. Esta recomendación es

evidentemente una apreciación subjetiva y hay que recordar que este tipo de evaluaciones fueron las que apoyaron el uso del clofibrato, la ligadura de la mamaria interna y la congelación gástrica. La necesidad de realizar una intervención terapéutica sin contar con suficiente información es una situación frecuente para los clínicos y la mejor forma de abordarla es considerando la causa y las consecuencias de los cursos alternativos de acción (ver la IV parte de esta serie): ¿Necesita el paciente de alguna intervención? Si así es, ¿existe algún tratamiento que haya demostrado ser más benéfico que perjudicial en un estudio aleatorio? Si no es así, ¿cuál de los tratamientos disponibles es el que parece ser la mejor alternativa al considerar beneficios y riesgo? Las siguientes guías pueden ser útiles para la evaluación de estudios clínicos aleatorios.

2. ¿Fueron informados todos los sucesos clínicos relevantes?

En la tabla 3, se muestran en forma resumida los resultados de un importante ensayo clínico aleatorio usando clofibrato en pacientes con niveles elevados de colesterol sérico (2).

Algunos de los resultados del tratamiento parecen ser muy favorables, por ejemplo, el nivel de colesterol sérico —un factor de riesgo— para enfermedad coronaria se redujo casi en un 10%, proporcionando evidencia biológica de beneficio. Sin embargo, algunos lectores reconocerán que el beneficio terapéutico que podría sustentarse con base en este cambio en los niveles de colesterol es un ejemplo del "juego de sustitución" en el que un factor de riesgo es substituido por la evolución clínica que se asocia con este factor (9). Estos lectores desearán saber si ocurrió un cambio real en la frecuencia de infartos coronarios.

Estos datos son también presentados en la tabla 3 en donde se observa una reducción en el número de infartos de miocardio no mortales y en el total de infartos (mortales y no mortales). Por lo tanto la eficacia del clofibrato parece demostrada en este

TABLA 3

SUCESOS CLÍNICOS RELEVANTES EN UN ENSAYO CLÍNICO ALEATORIO DE CLOFIBRATO PARA PREVENIR ENFERMEDAD CARDIACA CORONARIA*

| <i>Variable</i> | <i>Resultados</i> | |
|---|----------------------|-----------------------------|
| | <i>Grupo placebo</i> | <i>Grupo con clofibrato</i> |
| <i>Cambio promedio (%) en el nivel de colesterol sérico</i> | +1 | -9 |
| <i>Número de infartos miocárdicos no fatales/1000 hombres</i> | 7.2 | 5.8 |
| <i>Número de infartos miocárdicos fatales y no fatales/1000 hombres</i> | 8.9 | 7.4 |
| <i>Número de muertes/1,000 hombres</i> | 5.2 | 6.2 |

* Adaptado de la referencia 2.

estudio. Sin embargo, cuando consideramos todos los desenlaces de importancia clínica, especialmente desde el punto de vista del paciente, se deberán analizar los efectos del clofibrato sobre la calidad de vida y la mortalidad total, que también son mostradas de manera clara e inquietante en la misma tabla 3: la mortalidad fue mayor en el grupo tratado con clofibrato. Estos resultados afectaron mucho el uso y la disponibilidad de este medicamento en los E. U. A. Por lo tanto, en vista de ello se forma un juicio acerca de la utilidad del clofibrato o de otros agentes en función de los resultados clínicos seleccionados por los autores para comparación, el lector deberá estar convencido de que todos los resultados clínicos relevantes fueron informados. Además, dado que la discrepancia clínica es ubicua en medicina, los lectores deberán también reconocer que son necesarios criterios explícitos y objetivos para la evaluación de resultados clínicos de interés, criterios que deberán ser aplicados por observadores "ciegos", esto es, que ignoren que pacientes se encuentran en el grupo control y que pacientes reciben la droga en estudio.

3. ¿Fueron los pacientes estudiados similares a los que usted atiende?

Este punto tiene dos aspectos. Primeramente los pacientes deberán ser identificables; esto es, sus características clínicas y sociodemográficas deberán estar descritas con suficiente detalle para que le sea posible reconocer las semejanzas entre los pacientes estudiados y los pacientes de su consulta. En segundo lugar los pacientes del estudio deben ser *similares* a los pacientes de su consulta. Dicho de otra manera ¿Son los pacientes del estudio *tan diferentes* de mis propios pacientes que *no* pueda extrapolar los resultados a mi práctica? Este requisito va más allá de la cuarta guía general para la lectura de artículos clínicos (el lugar) y abarca las características precisas de pacientes individuales más que las características generales de los sitios de referencia.

TABLA 4

TRATAMIENTO MÉDICO Vs QUIRÚRGICO
EN ESTENOSIS CAROTÍDEA BILATERAL
(Resultados de todos los pacientes "disponibles
para seguimiento") *

Ataques de isquemia cerebral transitoria, accidente cerebro-vascular o muerte

| Tratamiento | Si | No | Total pacientes |
|---------------------|----|----|-----------------|
| Quirúrgico + Médico | 43 | 36 | 79 |
| | 53 | 19 | 72 |

* Adaptado de la referencia 14

+ Reducción del riesgo en tratamiento quirúrgico
(53/72) — (43/79) / (53/72) = 27%

$X^2 = 5.98$ y $p = 0.02$

Cuando estos dos aspectos, identificabilidad y similitud, son satisfechos, el lector puede suponer con confianza que obtendrá resultados similares a los descritos en su práctica.

4. ¿Se consideraron tanto la significación clínica como la estadística?

Aquí la significancia *clínica* se refiere a la *importancia* de encontrar una diferencia en la evolución de los pacientes tratados y los pacientes de grupo control y usualmente se describe en términos de *magnitud* de un resultado. En la tabla 3 se observa que los pacientes que recibieron clofibrato tuvieron una probabilidad 19% (6.2-5.2/5.2) mayor de morir que los pacientes que aleatoriamente recibieron placebo. Esta diferencia se hace clínicamente significativa cuando ocasiona cambios en el comportamiento clínico*, de esta manera, se confirma que esta diferencia de 19% en mortalidad total es clínicamente significativa cuando su reconocimiento motivó una drástica reducción en las prescripciones de clofibrato para estos pacientes.

En contraste, la significancia *estadística* únicamente nos dice si una diferencia es *real* y no si es importante o de gran magnitud. Más precisamente, el significado estadístico de una diferencia no consiste más que en establecer qué tan posible es que la diferencia sea debida a la casualidad únicamente. Por lo tanto, si la diferencia del 19% en mortalidad total entre los pacientes que recibieron clofibrato y aquellos que recibieron placebo tiene una posibilidad muy baja (por ejemplo: menos del 5% o < 0.05) de ser secundaria a la casualidad, se refiere a ésta como una diferencia estadística significativa.

Los determinantes de la significancia clínica son, por lo tanto, los determinantes de cambios en acciones clínicas; si los resultados de un estudio lo llevan a usted a abandonar un antiguo tratamiento por uno nuevo, la diferencia en los efectos de estos tratamientos es clínicamente significativa. Los determinantes de la significancia estadística no son tan obvios. En forma simple, la significancia estadística de cualquier resultado se eleva (esto es, el valor de P disminuye) cuando el número de pacientes en el estudio aumenta, cuando los efectos clínicos del tratamiento muestran menos fluctuaciones día a día o de paciente a paciente, y cuando la medida de este efecto clínico es precisa y reproducible.

Con base en lo anterior, el lector muy ocupado puede usar dos puntos de referencia para la lectura de artículos sobre terapéutica. Primero, si la diferencia es estadísticamente significativa ($P < 0.05$), ¿es también

* Aunque hemos definido significancia clínica desde el punto de vista del clínico, también podría definirse desde el punto de vista del paciente en términos de; "diferencias significativas en la calidad de vida".

clínicamente significativa? Si es así los resultados son tanto reales como dignos de llevarse a cabo en la práctica clínica. En segundo lugar, en caso de que la diferencia no sea estadísticamente significativa: ¿Son los pacientes suficientes en número para mostrar una diferencia clínicamente significativa si ésta hubiera ocurrido? Como se ha discutido previamente, el número de pacientes en un estudio es una de las determinantes de significancia estadística. Por lo tanto, si en un estudio la población es muy numerosa, las diferencias en su evolución clínica pueden ser estadísticamente significativas (reales) aún si clínicamente sean triviales (tan pequeñas que no justifiquen un cambio en el comportamiento clínico). A la inversa, si en un estudio la población es muy pequeña aún grandes diferencias de enorme potencial y significación clínica pueden no ser estadísticamente significativas.* Los lectores por lo tanto deben escudriñar las diferencias en la evolución clínica de los estudios cuyos resultados no son estadísticamente significativos para determinar si existe una potencial significancia clínica. Esta recomendación ha recibido apoyo adicional a partir de la demostración de que la mayoría de los ensayos clínicos aleatorios recientemente publicados cuyos resultados no fueron estadísticamente significativos tenían muy pocos pacientes para demostrar reducciones en riesgo de 25% o aún de 50% (12).

5. ¿Es el tratamiento propuesto utilizable en su práctica clínica?

En este punto existen cuatro requerimientos. Primero, la maniobra terapéutica tiene que ser descrita con suficiente detalle para que los lectores la puedan repetir con precisión. Quién le hizo qué a quiénes, qué presentación y en qué dosis, en qué circunstancias se administró, cómo se ajustó la dosis, qué tipo de respuestas y qué investigaciones se efectuaron para evaluar los efectos colaterales y la toxicidad, por cuánto tiempo y con qué criterios clínicos se decidió aumentar, disminuir o terminar el tratamiento.

En segundo término la maniobra terapéutica debe ser clínica y biológicamente razonable. Por ejemplo, la dosis, la vía de administración, y la duración del tratamiento deben ser consistentes con el conocimiento existente acerca de la farmacodinamia y de la farmacocinética del medicamento. Igualmente las combinaciones de diferentes tratamientos deben ser clínicamente razonables.

El tercer punto se refiere a la disponibilidad de

* Por convención esta posibilidad es llamada el "valor P", "alfa" o "la posibilidad de cometer un error tipo I" en las cuales se concluye que una diferencia existe cuando en realidad no es así.

la maniobra terapéutica. El médico lector deberá estar en posibilidad de administrar el tratamiento apropiadamente y sus pacientes deberán encontrar el tratamiento accesible, aceptable y costeable.

En cuarto lugar, cuando se lee la descripción de la maniobra se debe considerar si los autores evitaron 2 sesgos específicos en su aplicación: La contaminación, en la cual los pacientes del grupo control reciben accidentalmente el tratamiento experimental, lo que resulta en una reducción espuria en la diferencia de los resultados entre los grupos experimental y control; y la *cointervención*, cuando acciones diagnósticas o terapéuticas adicionales son efectuadas en los pacientes del grupo experimental pero no en el grupo control, lo que resulta en un aumento, también espurio, en la diferencia de resultados entre ambos grupos.

Una vez más resulta claro que la cointervención se previene en un estudio "ciego" en el cual tanto los pacientes como los médicos ignoren el tratamiento recibido.*

6. ¿Todos los pacientes que entraron al estudio fueron evaluados al finalizarlo?

El lector sagaz notará cuántos pacientes entraron al estudio (usualmente el número de pacientes en el grupo control y en el experimental serán casi idénticos) y llevará la cuenta de este número al término del estudio para estar seguro de que los números corresponden. Por ejemplo, en la tabla 4 se describe la evolución clínica de 151 pacientes en un estudio aleatorio de tratamiento médico vs. quirúrgico para la estenosis carótida bilateral (14). De 79 pacientes que se sometieron a tratamiento quirúrgico y 72 pacientes que recibieron tratamiento médico y que estuvieron "disponibles para seguimiento" (total de pacientes al final del estudio, 151) se encontró un 27% ($P=0.02$) de reducción en el riesgo de ataques de isquemia transitoria, de accidentes cerebro vasculares (ACV) o de muerte en aquellos que se trataron quirúrgicamente, con una diferencia que es clínica y estadísticamente significativa. Sin embargo, una cuidadosa lectura del informe muestra que fueron 167 los pacientes que iniciaron el estudio, no 151, y que 16 de ellos sufrieron un ACV o murieron durante su hospitalización inicial y fueron excluidos del análisis final. Aún más, 15 de estos¹ 16 pacientes fueron asignados a recibir tratamiento quirúrgico; 5 de ellos murieron y 10 tuvieron un ACV durante o poco después de la cirugía. El resultado de introducir estos 16 casos al análisis final se muestra en la tabla 5; la reducción

* Esto es a lo que se denomina "bajo poder", el error beta o el riesgo de un error de tipo II, en los cuales se concluye que una diferencia no existe, cuando en realidad sí existe.

TABLA 5

TRATAMIENTO MEDICO Vs. QUIRÚRGICO
EN ESTENOSIS CAROTÍDEA BILATERAL
(Resultados de todos los pacientes que fueron asignados
de manera aleatoria*)

| <i>Ataques de isquemia transitoria, accidente cerebro-vascular o muerte</i> | | | |
|---|-----------|-----------|------------------------|
| <i>Tratamiento</i> | <i>Sí</i> | <i>No</i> | <i>Total pacientes</i> |
| Quirúrgico + Médico | 58 54 | 36 19 | 94 73 |

* Adaptado de la referencia 14.

+ Reducción del riesgo en tratamiento quirúrgico:
(54/73) — (58/94) / (54/73) = 16%, $X^2=2.80$ y $P=0.09$

en riesgo con la cirugía es ahora únicamente del 16% y ya no es estadísticamente significativa ($P=0.09$.)

Los autores de este informe tuvieron el cuidado de incluir la evolución clínica de todos los pacientes que entraron al estudio, haciendo posible la elaboración e interpretación de la tabla 5.

¿Qué puede hacer el lector en el caso de que no se informe de la evolución de los pacientes perdidos?

Un abordaje (de antemano reconocido como conservador y en consecuencia sujeto a un error de tipo II) es el de asignar arbitrariamente una mala evolución a todos los pacientes "perdidos" del grupo que tiene los resultados más favorables.

Si esta maniobra no modifica la significancia clínica o estadística de los resultados a través de un punto de decisión, el lector podrá aceptar las conclusiones del estudio.

Uso de estas guías de lectura.

El abordaje a las publicaciones clínicas descrito en esto y en los otros artículos de esta serie sobre Epidemiología Clínica ha sido diseñada para los clínicos atareados que desean mantenerse actualizados en los avances más importantes en diagnóstico clínico, en nuevos cauces sobre el curso clínico y el pronóstico de las enfermedades, en descubrimientos acerca de la etiología de la enfermedad y en la existencia de mejorías clínicamente significativas en terapéutica. Los clínicos enfrentamos una tarea agobiante aunque ya atrasados en nuestra lectura clínica, se nos pide que observemos el contenido de un número cada vez mayor de revistas.

Asumiendo que nunca tendremos más tiempo para leer que el que tenemos ahora y reconociendo que se requiere de una evaluación crítica de la literatura clínica, si es que pretendemos beneficiar y no perjudicar a nuestros pacientes, hemos organizado esta serie de guías de sentido común para evaluar los artículos clínicos. Uno de los resultados principales de su aplicación es evitar la lectura de muchos, sino es que la mayoría de los artículos clínicos.

Indudablemente que dejarán de leerse algunos buenos artículos al aplicar estas reglas, pero creemos que aquellos artículos que "sobrevivan" después del análisis sugerido serán los más válidos, los más relevantes y los más aplicables a nuestra práctica clínica.

Conclusión

¿Qué proporción de artículos llenarán los requerimientos de aplicabilidad clínica y valor científico descrito en estas revisiones de Epidemiología Clínica? Evidentemente, no muchos, aunque hay evidencia de que la situación está mejorando.*

Después de todo, hay apenas unos cuantos caminos para llevar a cabo un estudio de manera adecuada y hay miles de maneras para llevarlo a cabo incorrectamente. Más aún, si un estudio satisface todos los requerimientos, esto solucionará una situación clínica para todo el tiempo. En el mejor de los casos contribuirá un poco, algunas veces sólo temporalmente, a aumentar nuestra capacidad para aliviar el sufrimiento y promover la salud. Asimismo, aún los estudios mejor diseñados provocarán en ocasiones profundas y prolongadas controversias.

Las razones para la lentitud del progreso y para las disputas son varias. En primer lugar, existe la posibilidad de que a pesar de un diseño y un análisis impecables, los resultados del estudio sean incorrectos; esta es una consecuencia inevitable, aunque rara, de probar los resultados por su significancia estadística: ocasionalmente los resultados serán debidos únicamente al azar.

Segundo, el entendimiento contemporáneo de la estructura humana, su funcionamiento y los mecanismos de enfermedad que llevan a los investigadores clínicos a agrupar a ciertos pacientes o respuestas podrían subsecuentemente encontrarse gravemente deficientes contradiciendo así los resultados o las interpretaciones del estudio original.

Tercero, un estudio puede ser malentendido o malinterpretado por quienes lo leen, como ocurre cuando un estudio diseñado para resolver una pregunta como: ¿Podría el tratamiento X funcionar en circunstancias óptimas (por ejemplo: pacientes apegados al tratamiento, esquemas de dosis muy elaboradas y un restringido espectro de evoluciones clínicas) ?, y se le critica por su incapacidad para responder preguntas de manejo, "¿es el tratamiento X benéfico en circunstancias clínicas habituales (por ejemplo: cualquier paciente, esquema de dosis habituales y toda una gama de evoluciones clínicas)" (10).

* Aunque en revistas médicas generales los estudios de cohortes parecen menos frecuentes que los estudios transversales, que tienen menos poder, los estudios clínicos aleatorios

Cuarto, puede existir controversia en la interpretación inclusive de estudios válidos cuando se analizan objetivos diferentes a los planteados por los autores. Por ejemplo, los estudios de tratamientos alternativos para manejar pacientes con síntomas de apendicitis han demostrado que se puede minimizar el número de muertes por esta condición con una política quirúrgica liberal en la que todos los pacientes aún aquellos con síntomas discretos se operen (16).

Por otra parte si uno quiere minimizar el número de cirugías innecesarias, costos de hospital o el tiempo de convalecencia, uno podría adoptar una política más conservadora y limitarla a los pacientes con síntomas graves. En estos casos hay dos "mejores respuestas" para una misma situación clínica y la controversia se hace inevitable.

Quinto, los resultados e interpretación de estudios, aún de aquéllos que satisfacen los requerimientos asentados en estos artículos, pueden enfrentar considerable resistencia cuando desacreditan al único abordaje terapéutico clínicamente disponible para alguna condición, puesto que los clínicos a pesar del estudio pueden elegir hacer algo, aún cuando este algo no haya resultado benéfico, en lugar de hacer nada. Finalmente, los resultados pueden ser rechazados, independientemente de su mérito, si amenazan el prestigio o la sobrevivencia de su público.

En resumen, esta serie de discusiones intenta ayudar al lector responsable a evaluar apropiadamente la literatura clínica que potencialmente ofrece nuevos conocimientos válidos y útiles en menos tiempo. Aunque sería ingenuo de nuestra parte esperar que la aplicación de estas reglas acelerará grandemente la adquisición y la aplicación clínica de nuevos conocimientos si confiamos en que su adopción asegurará que la inercia que se adquiera sea hacia adelante.

Aunque estas guías para el lector se han presentado para su uso en la lectura de la literatura clínica actual puede también tener otros usos. Por ejemplo, pueden ser útiles en una revisión de literatura enfocada hacia la identificación de los artículos potencialmente más útiles. Más aún se pueden aplicar al pie de la cama del enfermo o durante el desarrollo de las sesiones clínicas o bibliográficas que versen sobre diagnóstico, pronóstico, etiología o tratamiento. Finalmente, se pueden usar para organizar y presentar evidencia sobre diagnóstico, pronóstico, etiología y tratamiento a estudiantes y colegas. Será bienvenida la retroalimentación sobre la utilidad de esta serie para todos estos propósitos así como sugerencias para su mejora.

AGRADECIMIENTO

Agradecemos a nuestros estudiantes, residentes y colegas clínicos, por sus sugerencias y críticas de versiones iniciales de este trabajo.

REFERENCIAS

1. Glover RP, Davila JC, Kyle RH, Beard JC Jr, Trout RG & Kitchell JR: Ligation of the internal mammary arteries as a means of increasing blood supply to the myocardium. *J Thorac Cardiovasc Surg*, 34: 661, 1957.
2. Oliver MF, Heady JA, Morris JN & Cooper J: A co-operative trial in the primary prevention of ischaemic heart disease using clofibrate. Report from the committee of principal investigators.
1. Glover RP, Davila JC, Kyle RH, Beard JC Jr, Trout RG & Kitchell JR: Ligation of the internal mammary arteries as a means of increasing blood supply to the myocardium. *J Thorac Cardiovasc Surg*, 34: 661, 1957.
2. Oliver MF, Heady JA, Morris JN & Cooper J: A co-operative trial in the primary prevention of ischaemic heart disease using clofibrate. Report from the committee of principal investigators. *Br Heart J*, 40: 1069, 1978.
3. Miao LL: Gastric freezing: an example of the evaluation of medical therapy by randomized clinical trials. En Bunker JP, Barnes BA, Mosteller F (Eds): *Costs, Risks and Benefits of Therapy*, Oxford U Pr, New York, 198-211: 1977.
4. Cobb LA, Thomas GI, Dillard DH, Merendino KA & Bruce RA: An evaluation of internal-mammary-artery ligation by a double-blind technic. *N Engl J Med*, 260: 1115, 1959.
5. Bywaters EG: Treatment of rheumatic fever. *Circulation*, 14: 1153, 1956.
6. Sinclair JC: Prevention and treatment of the respiratory distress syndrome. *Pediatr Clin North Am*, 13: 711, 1966.
7. Hinshaw HC, Feldman WH & Pfuetze KH: Treatment of tuberculosis with streptomycin: summary of observation on 100 cases. *JAMA*, 132: 778, 1946.
8. Medical Research Council: Streptomycin treatment of tuberculous meningitis. *Lancet*, 1: 582, 1948.
9. Yerushalmy J: On inferring causality from observed associations. En Ingelfinger FJ, Relman AS, Finland M. (Eds): *Controversy in Internal Medicine*, Saunders, Philadelphia, 1966: 659-668.
10. Sackett DL & Gent M: Controversy in counting and attributing events in clinical trials. *N Engl J Med*, 301: 1410, 1979.
11. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont: Clinical disagreement: I. How often it occurs and why. *Can Med Assoc J*, 123: 499, 1980.
12. Freiman JA, Chalmers TC, Smith H Jr & Kuebler RR: The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. *N Engl J Med*, 299: 690, 1978.
13. Sackett DL: Design, measurement and analysis in clinical trials. En Hirsh J, Cade JF, Gallus AS, Schonbaum E. (Eds): *Platelets, drugs and thrombosis: Memories of a Symposium* llevado a cabo en la Universidad McMaster, Hamilton, Ont, Octubre 16-18, 1972, Karger, Basel, 219-225: 1975.
14. Fields WS, Maslenikov V, Meyer JS, Hass WK, Remington RD & Macdonald M: Joint study of extracranial arterial occlusion. V. Progress report of prognosis following surgery or nonsurgical treatment for transient ischemic attacks and cervical carotid artery lesions. *JAMA*, 211: 1993, 1970.
15. Fletcher RH & Fletcher SW: Clinical research in general medical journals. A 30-year perspective. *N Engl J Med*, 301: 180, 1979.
16. Neutra R: Indications for the surgical treatment of suspected acute appendicitis: a cost-effectiveness approach. En Bunker JP, Barnes BA, Mosteller F. (Eds): *Costs, Risks and Benefits of Surgery*, Oxford U. Pr, New York, 277-307: 1977.