



# *Apreciación crítica de la literatura médica*

**Gabriel Gómez, Hernando Gaitán, Rodrigo Pardo**

El siguiente es el resumen del simposio que sobre el tema se llevó a cabo en el VII Congreso de la Asociación de Exalumnos de Medicina de la Universidad Nacional de Colombia (Aexmun). El artículo consta de tres partes: la primera introductoria escrita por G. Gómez sobre cómo y por qué evaluar críticamente la literatura biomédica, basado en los artículos y en el libro que sobre el tema han publicado docentes de la Universidad McMaster (1,2), la segunda sobre el análisis de exámenes diagnósticos por H. Gaitán, y la tercera sobre análisis de estudios sobre tratamientos, por R. Pardo.

## **Introducción**

Los cambios acelerados que están teniendo lugar en la práctica médica, las labores asistenciales a las cuales debe dedicar gran parte de su tiempo el profesional, el cada vez mejor y más fácil acceso a la información por vía de la sistematización, el lenguaje en la producción de evidencia científica diferente y especializado con relación a la labor cotidiana, son factores que unidos a la evolución vertiginosa del conocimiento médico (consideremos que el compendio de éste se modifica totalmente cada 10 años) hacen que la autogestión continuada en la educación médica sea el pilar y requisito fundamental para una práctica de calidad.

Conocer el lenguaje de la ciencia médica no es suficiente para interpretar los trabajos de investigación; se hace indispensable, además, acceder a la terminología metodológica de la investigación en clínica y a los principios que la rigen; sólo así facilitaremos la puesta en marcha de una solución práctica, oportuna y eficiente a las situaciones, de nuestros pacientes. Es entonces el movimiento de un contexto individual de la práctica a un contexto más universal.

Hoy, por lo tanto, debemos reclamar una práctica médica basada en "evidencia científica" que vaya más allá de la simple experiencia clínica y de los está-

ticos principios fisiológicos. Se propone, por lo tanto, que el profesional en la actualidad esté en la capacidad de hacer evaluaciones estrictas de todas sus acciones clínicas y proyectarlas en las posibles consecuencias. Esto es tener la capacidad de moverse entre los lenguajes que les competen, que por un lado es el propio o cotidiano de la profesión y por el otro es el contexto universal del método científico.

La revisión que se pretende hacer en esta presentación no es para la aplicación en el diseño de una investigación, sino para saber interpretarla.

Las fuentes principales de la información médica son usualmente:

- La experiencia con pacientes individuales
- Resultados de investigaciones
- La consulta con expertos (la cual está sujeta a las interpretaciones subjetivas del consultado).
- Los libros, que presentan un conjunto organizado temáticamente para el consultante, igualmente viciada por la interpretación del lector. Debemos anotar que esta fuente está particular-

Dr. Gabriel Gómez Morales: Profesor Asistente de Epidemiología Clínica y Cirugía, Facultad de Medicina, Pontificia Universidad Javeriana; Dr. Hernando Gaitán: Profesor Asistente de Ginecología, Universidad Nacional de Colombia, Programa de Maestría en Epidemiología Clínica, Pontificia Universidad Javeriana, Dr. Rodrigo Pardo: Profesor Asistente de Neurología, Universidad Nacional de Colombia, Programa de Maestría en Epidemiología Clínica, Pontificia Universidad Javeriana

## Evaluación de la literatura

mente sometida a rápida desactualización.

Los archivos personales pueden enunciarse como otra fuente de información, la cual debe nutrirse permanentemente, pues corre el riesgo de perder vigencia.

• Finalmente, la consulta a bases de datos, de trabajos científicos, se considera como la fuente más eficiente por su permanente actualización, facilidad de consulta y por la información de primera mano (no procesada por intermediarios), pero primordialmente por la posibilidad que ofrece al consultante de ejercer sobre ella una revisión sistemática y crítica. Por el carácter conceptual de la metodología científica, se impone la necesidad de un sano criterio que racionalice el enfoque del conjunto informativo que procede de esta última fuente de información. Es la oportunidad que tiene el lector de hacerse a una idea personal relacionándose directamente con los trabajos de los investigadores.

La problemática del médico se origina desde el momento en que debe adecuar la situación particular que presenta su paciente, a la descripción genérica y abstracta contenida en las fuentes de información.

En esta materia el trabajo del profesional en salud consiste en trasladar situaciones prácticas a enunciados que se formulan como preguntas, las cuales deben responderse consistentemente, verificándolas cuando sea factible (con base en los resultados de investigaciones precedentes). Una vez identificadas las preguntas deben ser codificadas en el lenguaje universal de las bases de datos. La variabilidad y multiplicidad de los eventos para una única pregunta hace de la medicina una ciencia

no exacta que impone una aguda percepción en la interpretación de las probabilidades que ofrecen los artículos.

Una vez ejecutado el paso de la formulación de la pregunta, y encontrada su descripción pertinente, ésta debe ser contextualizada en la realidad de su virtual aplicación.

El paso siguiente es al revisión de la metodología empleada en cada estudio, descartando de primera mano los que no se adecúan a sus propósitos. Sobre este punto en particular señalaremos los requisitos metodológicos que hacen aceptables o no las hipótesis verificadas o rechazadas por los autores.

Los trabajos de investigación se pueden clasificar de acuerdo con sus características metodológicas en: estudios de tratamiento, de diagnóstico, de pronóstico, y de causalidad.

### Estudios de tratamiento

Ubiquemos en el escenario clínico e imaginemos el siguiente caso: un hombre de 65 años con hipertensión arterial controlada y con hipertensión y con historia de fibrilación auricular de seis meses de evolución, resistente a cardioversión. El médico que remite al enfermo ha descartado enfermedad valvular y coronaria y está interesado en saber el beneficio del uso de anticoagulantes a largo plazo (para reducir el riesgo de embolismo). Además el paciente está preocupado por la posibilidad de sangrado con la droga formulada.

Lo primero que debe se hacer es buscar los artículos que pudiesen tener la respuesta de los beneficios del tratamiento y sus probables complicaciones. Para

la evaluación de éstos se deben conocer los siguientes puntos:

#### 1. Forma de presentación de los resultados.

##### a. Eficacia del tratamiento

- Riesgo atribuible.
  - Riesgo relativo
  - Reducción del riesgo relativo
- ##### b. Precisión para estudios de tratamiento
- Intervalo de confianza de 95%.
  - Valor de p.
  - Tamaño de la muestra
  - Error estándar
  - Siempre comparación de los beneficios contra los riesgos y los costos de los tratamientos suministrados.

#### 2. Validez

- Criterios de inclusión y exclusión
- Seguimiento adecuado de la población y de los eventos.
- No contaminación
- Comparación ciega
- Homogeneidad de los grupos de comparación, (aleatorización).

#### 3. Aplicación de los resultados

Traer los resultados a la realidad del paciente en particular.

- ¿Todos los resultados pertinentes fueron considerados?
- Los beneficios de este tratamiento superan a los potenciales efectos adversos peligrosos y costosos.

### Estudios de pruebas diagnósticas

En un paciente de 75 años de edad en postoperatorio de cirugía abdominal se sospechó tromboembolismo pulmonar, por lo cual se le realizó una gamagrafía de ventilación-perfusión, que dio una probabilidad intermedia del evento. ¿Qué debe hacer el mé-

dico? ¿Instaurar terapia formal de anticoagulación?

### 1. Resultados

¿Se presentan los resultados con probabilidades? Sensibilidad, especificidad, valores predictivos, prevalencia.

¿Se incluyen datos para poder recalculan los presentados?

### 2. Validez

¿La comparación fue independiente y ciega con respecto al patrón de referencia?

¿La población corresponde al tipo de paciente a quien se le aplicará el estudio?

El resultado del estudio evaluado tiene influencia en la aplicación del patrón de oro?

- ¿El estudio puede ser reproducido?

### 3. Aplicación

- ¿Son reproducibles los probables resultados en el contexto de mis pacientes?

- ¿Los resultados cambian la conducta con los pacientes?

- ¿La aplicación de los resultados redundan en beneficio para el paciente?

## Estudios de pronóstico

Una paciente de 42 años de edad llega a consulta después de una mastectomía radical modificada y del primer ciclo de quimioterapia y pregunta cuál es la posibilidad de recidiva del tumor durante el primer año después del tratamiento

### 1. Resultados

- ¿Magnitud de la probabilidad de presentación del evento?

- ¿Qué tan preciso fue el estimativo de la probabilidad?

### 2. Validez

- ¿La muestra de pacientes fue bien definida, es representativa

y tiene un punto similar en el curso de la enfermedad?

- ¿El seguimiento fue suficiente y completo?

- ¿Se definieron objetivos y resultados previendo posibles sesgos?

- ¿Se ajustó y estratificó por aquellas variables que pudieran haber modificado el resultado?

### 3. Aplicación

- ¿La población es similar a la de mis pacientes?

- ¿Los resultados llevan a modificar o evitar terapias?

- ¿Los resultados sirven como información para el médico y los pacientes?

## Artículos de causalidad

En un artículo real se llegó a la conclusión que el consumo de café era un factor de riesgo para el cáncer de páncreas con todas las estadísticas reafirmando estos resultados.

### 1. Resultados

- ¿Qué tan consistente es la asociación entre la exposición y el resultado?

- ¿Qué tan precisa es la estimación del riesgo?

### 2. Validez

- ¿Son los grupos comparables y homogéneos?

- ¿La medición de la exposición y de los resultados fue adecuada?

- ¿El seguimiento fue suficiente y completo?

- ¿Hay relación temporal entre la exposición y el resultado a ésta?

- ¿Hay un gradiente entre la dosis y la respuesta?

### 3. Aplicación

- ¿Son los resultados aplicables a mi práctica?

- ¿Cuál es la magnitud del riesgo?

- ¿Se podría controlar la exposición?

## Conclusiones

La revisión crítica y sistemática de los artículos médicos debe tener como consecuencia el mejoramiento del conocimiento, la proyección de los eventos observados en la práctica clínica diaria y la anticipación de sus posibles soluciones y/o consecuencias.

El médico en forma permanente debe evaluar sus acciones y las consecuencias de éstas en los pacientes, cuidando de hacer lo mejor y no de ponerlos en riesgo. Esta evaluación tiene que ser objetiva y contextualizada en parámetros más universales.

Las decisiones basadas en análisis superficiales pueden llevar a conductas erróneas con costos humanos y económicos elevados.

## LECTURA CRITICA DE LA LITERATURA EN EXAMENES DIAGNOSTICOS

La intención de la presente revisión es dar algunas recomendaciones sobre los pasos por seguir en la evaluación de los artículos de revistas o libros que consideran el tema del diagnóstico, sea basado en hallazgos clínicos, de laboratorio o de imágenes diagnósticas tan de moda en la actualidad. Comenzaremos haciendo algunas consideraciones generales, luego enumeraremos los pasos por seguir en la evaluación dando algunos ejemplos para un mejor entendimiento de los tópicos, los cuales por razones de la especialidad del autor estarán más en el campo de la ginecología y obstetricia, pero salpicados por ejemplos clásicos de otras especialidades. La

revisión se hizo con base en los textos de Epidemiología Clínica de Sacket (2), Fletcher (3) y Moreno Altamirano (4), además de los artículos que sirven para los ejemplos.

El examen clínico es mucho más poderoso que la mayoría de las pruebas de laboratorio para establecer el diagnóstico, pronóstico y tratamiento en la mayoría de los pacientes. Las imágenes diagnósticas y otras pruebas de laboratorio han mejorado nuestra capacidad diagnóstica al proveer "datos precisos"; sin embargo, los datos obtenidos del paciente por medio de la historia clínica siguen siendo muy útiles: por ejemplo, en el paciente con angina de pecho típica, podrá tenerse una certeza muy alta de que tenga estenosis de 75% de las arterias coronarias, en este caso la arteriografía se hace más por las implicaciones terapéuticas (revascularización) que por el aporte al diagnóstico en sí.

Para mejorar el diagnóstico clínico deberíamos tener mediciones clínicas más confiables y reproducibles y evitar aquellas mal recogidas y sesgadas que conducen a diagnósticos incorrectos, cuyas consecuencias pueden ser devastadoras como sucede en nuestro medio con el mal diagnóstico en los servicios de urgencias, frecuentemente llenos de pacientes y con escaso personal.

Acerca del problema de la reproducibilidad de los hallazgos clínicos debemos tener en cuenta la variabilidad entre los observadores o inclusive en un observador al interpretar un mismo examen, transcurrido un determinado tiempo. Esta situación lleva a que puede darse una discrepancia entre las dos observa-

ciones, lo cual es de común ocurrencia en la historia clínica, en los hallazgos físicos o en la interpretación de pruebas diagnósticas como radiografías, ecografías o colposcopias.

Pero, ¿para qué reunimos datos diagnósticos? (clínicos, de laboratorio, imágenes). Las razones las podemos enumerar así:

a. Para hacer un diagnóstico preciso (preeclampsia).

b. Como medio para juzgar la gravedad de un síndrome en lugar o además de su causa. Así la presencia de alteraciones visuales, hiperreflexia y cefalea nos alertan sobre la gravedad de la preeclampsia

c. Para juzgar la evolución clínica y el pronóstico. Razón por la que buscamos las metástasis en pacientes con diagnóstico de cáncer.

d. Respuesta probable al tratamiento en el futuro. Así determinamos la presencia de receptores estrogénicos en el carcinoma de mama

e. Respuesta probable al tratamiento en el presente. Solicitando pruebas tiroideas después de la iniciación del tratamiento en la paciente con hipotiroidismo.

El criterio al evaluar una prueba diagnóstica debería ser la utilidad para quien la solicita y para el paciente. ¿Cómo hacer para buscar la evidencia de la utilidad de las pruebas, más aún en la avalancha de nuevos métodos diagnósticos?

Un método sería por medio de la actualización, es decir, la selección, evaluación, aprendizaje y archivo de la literatura, motivo de otro de los artículos de estas publicaciones, o la otra, discriminar los datos para decidir cuáles deben ser investigados, median-

te la evaluación de recomendaciones de "expertos" tal vez ideal en casos muy poco frecuentes o por medio de la apreciación crítica de los datos sobre métodos diagnósticos en los síndromes de frecuente presentación.

¿Cómo se puede decidir si un síntoma, signo o prueba diagnóstica es útil?

Analizaremos cada uno de los criterios sugeridos por Sacket:

1. ¿Se realizó una comparación a ciegas, independiente, con un patrón diagnóstico?

a. ¿Hubo un patrón de oro válido? Generalmente el patrón de oro es el examen que "tiene la verdad" sea anatomía patológica, para padecimientos como el cáncer, seguimiento prolongado en casos de enfermedades crónicas como la artritis reumatoidea u otro patrón reconocido que usted pueda aceptar, como sería la elección de las enzimas séricas y la alteración del electrocardiograma para el caso del infarto del miocardio. ¿Si el patrón no es claro o no reconocido, vale la pena continuar?

Además deberemos tener en cuenta que estos patrones pueden cambiar al mejorar el conocimiento de la biología de una entidad o la tecnología disponible para su estudio como sucedió con la colecistografía, la cual fue reemplazada por el ultrasonido para el diagnóstico de la coleditiasis.

Si el patrón es válido entonces considere si la prueba diagnóstica tiene algo que ofrecer que el patrón no tiene: más económico, menos molestias, se puede aplicar más tempranamente. Esto es muy válido para los datos clínicos de obtención inmediata. Si no ofrece ventajas, ¿se justifica continuar adelante?

Si se continúa adelante, entonces se entrará a evaluar la validez de la prueba, es decir, los niveles de sensibilidad, especificidad, valores predictivos y probabilidades, si parecen ser promisorios, continúe; si no es así y la prueba no es mejor, pregúntese nuevamente sobre sus beneficios adicionales,

b. ¿Fue realmente a ciegas la evaluación? Es decir, aquellos que están evaluando la prueba nueva, desconocían el verdadero estado del paciente. De la misma manera quienes están aplicando la prueba patrón, no deben saber el resultado de la otra prueba en ningún paciente; sólo así se podrán descartar sesgos de sospecha diagnóstica o de expectativa (5).

A continuación haremos una breve descripción de la manera como se evalúa la validez de una prueba: la determinación de la validez de un prueba en un padecimiento se establece por medio de la sensibilidad y especificidad de la misma con relación a dicha enfermedad. Son los índices que nos dan la eficacia de la prueba para establecer o descartar determinado diagnóstico, así el clínico ha de conocer dichos valores de las pruebas que utiliza diariamente para darles el peso adecuado en sus decisiones.

Al evaluar la eficacia de una prueba para generar un diagnóstico, se compara como dijimos con el examen (signo clínico, laboratorio, etc.) que se dice poseer la verdad o por lo menos estar más cercano a ella. Al realizar la prueba se obtendrán por lo menos dos resultados: positivo cuando se cree que presenta la enfermedad y negativo cuando se cree que no la presenta. Al compararla con la certeza diag-

nóstica se obtendrán dos resultados correctos y dos incorrectos: los correctos son verdadero positivo y verdadero negativo; y los incorrectos, falso positivo cuando dice que tiene la enfermedad y no la tiene realmente, y falso negativo cuando dice que no la tiene, pero sí hay la enfermedad.

Si se construye una tabla de contingencia de 2x2 tendríamos (Tabla 1).

**Sensibilidad:** probabilidad que una prueba esté positiva dado que el individuo esté enfermo p t(+)/D(+), también se le conoce como tasa de positividad en presencia de enfermedad (TPP), o tasa de verdaderos positivos (tasa **VP**).

**Especificidad:** posibilidad de que la prueba esté negativa dado que el individuo no tiene la enfermedad p t(-)/D(-), tasa negativo en ausencia de enfermedad (NAE), o tasa de verdaderos negativos (tasa **VN**). Para que una prueba sea buena se necesita que sea sensible y específica, es decir, que no sólo detecte casi todos los individuos enfermos, sino que detecte los sanos por los resultados negativos (2, 5).

La sensibilidad y especificidad de la prueba se aplican para una determinada enfermedad, siendo diferentes para otras entidades.

**Cálculo de la sensibilidad y especificidad:** de acuerdo con la tabla de contingencia descrita anteriormente la sensibilidad se determinará como la cantidad de enfermos, con la prueba positiva sobre el total de los enfermos es decir =  $a/a+c$ . Para la especificidad el numerador tendrá el número de personas con la prueba negativa sin la enfermedad y el denominador el número total de personas sin la enfermedad =  $d/b+d$ .

Enfermedad (dado por el patrón)		
	Presente	ausente
Prueba	Positiva VP (a)	FP (b)
	Negativa	FN (c) VN (d)
a + c: número total de enfermos		
b + d: número total de personas sin la enfermedad		
a + b: número total de personas con la prueba positiva		
c + d: número de personas con la prueba negativa		
a + b + c + d: número total de personas estudiadas		
a + c / a + b + c + d : prevalencia de la entidad		
a + d / a + b + c + d : utilidad global de la prueba		

**Tabla 1.** Tabla de 2x2 de la relación enfermedad y prueba diagnóstica.

Evaluación de la literatura

Malformación congénita (autopsia)			Total
Ultrasonido	Casos	Controles	
	(+) 152	2	154
	(-) 135	285	420
	287	287	574
Sensibilidad: 152/287 = 53%			
Especificidad: 285/289 = 99%			

Tabla 2. Relación entre enfermedad (malformación) y prueba (ultrasonido).

Enfermedad pélvica inflamatoria		
	(+)	(-)
(+) 23	3	26 VPP = a/a+b = 88%
secreción endocervical purulenta		
(-) 19	14	33 VPN = d/d+c = 42%
	42	17 59
sensibilidad a/a+c = 55%		especificidad d/d+b = 82%

Tabla 3. Relación entre enfermedad (EPI) y prueba (secreción purulenta), prevalencia alta.

Enfermedad pélvica inflamatoria		
	(+)	(-)
(+) 23	47	70 VPP = 32%
Secreción endocervical purulenta		
(-) 19	211	230 VPN = 91%
	42	258 300
sensibilidad a/a+c = 55%		especificidad d/d+b = 82%

Tabla 4. Relación entre EPI y prueba (secreción purulenta), prevalencia baja.

Si la sensibilidad es alta, cercana a uno, habrá pocos falsos negativos y se dirá que la prueba es "SENEX" (si negativo excluye). Si la especificidad es cercana a uno se dirá que hay pocos falsos positivos o sea que la prueba es "ESPIN" (es positivo incluye), se invoca ESPIN cuando se lleva a cabo el reconocimiento del patrón.

Ejemplo: Goncalvez y colaboradores en 1994 compararon la utilidad del ultrasonido en el diagnóstico de malformaciones fetales con los resultados del examen físico a la salida del recién nacido del hospital o por autopsia cuando hubo pérdida perinatal. Obtuvieron los siguientes resultados: (Tabla 2).

Aquí podemos concluir que si se detecta una alteración, se tiene una altísima probabilidad (99%) de que ésta se encuentre realmente (ESPIN). Cuando la prueba tiene casos dudosos como sucede frecuentemente en imágenes diagnósticas se pueden obtener resultados de dos tipos. Como primera medida se puede evidenciar una alteración, pero no es posible darle una interpretación precisa. Se clasificarán como casos dudosos con el padecimiento (Dp). De otra parte estudios practicados a personas enfermas cuyos resultados no aportaron suficientes datos para descartar la enfermedad, sin ser totalmente "normales", son casos dudosos sin el padecimiento (Ds). Las observaciones dudosas no se pueden eliminar al hacer una investigación en métodos diagnósticos y deben ser tenidas en cuenta.

Dependiendo del grado de alteración que el examen quiera detectar, la prueba podrá ser positiva, dudosa o negativa: como

por ejemplo el empleo de la resonancia nuclear para diferenciar entre absceso tuboovárico o tumor ovárico o la ecografía para diferenciar entre tumores malignos o benignos de ovarios

Estos resultados (Dp) se deberán incluir en el denominador de la sensibilidad o (Ds) en el denominador de la especificidad  $S = a/a+c+Dp$   $E = d/d+b+Ds$  (4). Confiabilidad de los índices de sensibilidad y especificidad: muy frecuentemente los índices de S y E se determinan en grupos de personas que tienen la enfermedad y en otros que no la tienen, extrapolándose las observaciones a la población blanco. Son proporciones en las que el numerador queda incluido en el denominador dado por el total de enfermos o de personas sin la enfermedad.

Además se aplican en poblaciones con una prevalencia dada, la cual puede variar de una población general donde si la muestra es representativa son más generalizables, o si son tomados por ejemplo de una institución de remisión donde son menos generalizables ya que la prevalencia aumenta, aunque esto tendrá más efectos sobre los valores predictivos.

Selección de la prueba con base en la sensibilidad: se busca sensibilidad alta cuando la enfermedad es de mal pronóstico y no debe pasar inadvertida, sobre todo si hay un tratamiento efectivo, es decir, nos interesaría disminuir los falsos negativos como sería el caso de la estenosis renal, o alteraciones neonatales como por ejemplo el hipotiroidismo congénito, en las cuales el punto de exclusión debe ser lo suficientemente bajo como para que la prueba identifique

virtual o absolutamente a todos los pacientes con la alteración blanco (no deseando perder ningún caso, no nos importa el aumento de los falsos positivos tratando de disminuir los falsos negativos).

Es deseable tener alta especificidad cuando la enfermedad es de mal pronóstico o incurable o cuando los falsos positivos puedan tener consecuencias graves para el paciente o la familia como sería en pacientes sometidos a una prueba para el SIDA, o para el cáncer cuando se desea disminuir los falsos positivos al mínimo, es decir, el punto de inclusión se tendría muy alto de modo que excluya al 99% de los sanos, teniendo muy pocos falsos positivos. Si se ha establecido ya el diagnóstico en un paciente, y se tienen dos pruebas para elegir se escogerá aquella que tenga la especificidad más alta (ESPIN).

La mayor ganancia entre la probabilidad antes y después de la prueba, se da con sensibilidad y especificidad de 80 a 90% y con prevalencias (probabilidad de tener la alteración antes de la prueba) entre 40 y 60%, sitio en el cual se logra un mayor beneficio para el paciente y que explica por qué mediante el llamado arte de la medicina o juicio clínico, nuestros profesores acertaban en el diagnóstico de entidades ubicándose sin saberlo muchas veces en este rango de probabilidades al cual le aplicaban pruebas clínicas con buena sensibilidad y especificidad mejorando la probabilidad de detectar la enfermedad después de la misma; esta probabilidad preprueba siempre será mayor entre más experiencia se tiene en la alteración blanco.

Las curvas COR (curva operativa para el receptor) nos dan el punto de corte donde se obtienen la mejor sensibilidad y especificidad (dado por  $1 - \text{especificidad}$ ) y se deduce por dónde es menor la suma de FP y FN o  $S+E/2$  y se relaciona con la importancia en el paciente de un falso positivo o falso negativo según lo explicado anteriormente. Orientan el punto de corte para una serie de valores de una prueba.

Sin embargo, en la práctica clínica no sabemos quién tiene en realidad la alteración blanco, si lo supiéramos no tendríamos que hacer la prueba. Nuestro interés es más bien la interpretación de los resultados positivos y negativos de la prueba o valores predictivos; la probabilidad de que dado que la prueba es positiva que el individuo tiene la enfermedad  $p D(+)/T(+)$  = VPP o valor predictivo positivo, o dado de que la prueba es negativa la probabilidad de que no tenga la enfermedad  $p D(-)/T(-)$  = VPN o valor predictivo negativo. Ambas varían directamente con la prevalencia y recordando que ésta representa la probabilidad antes de la prueba de tener el padecimiento, es todavía más importante el hecho de los datos obtenidos en una determinada población son sólo aplicables a ella.

Ejemplo: en el servicio de sépticas del Instituto Materno Infantil la prevalencia de enfermedad pélvica inflamatoria entre 1992-1993 en una muestra de 63 pacientes fue de 66% y en esta muestra la secreción purulenta endocervical como signo diagnóstico se obtuvo en 59 pacientes, observándose el siguiente comportamiento (7) (Tabla 3).

Tendríamos en esta situación una probabilidad de 88% de tener la enfermedad dado que hubiera secreción endocervical purulenta. Si se viera dentro del contexto general del servicio de sépticas del instituto con una prevalencia de 14%, con la misma sensibilidad y especificidad, tendríamos en una muestra de 300 pacientes: (Tabla 4).

Aquí la probabilidad de tener enfermedad pélvica inflamatoria sería de 32%. al observar la secreción endocervical purulenta. Así la disminución del VPP es el efecto más dramático de la disminución de la prevalencia. Se observa además un aumento del VPN, este aumento de los valores predictivos negativos es importante con prevalencias altas, pero muy pequeño en prevalencias bajas.

El máximo beneficio del valor predictivo de una prueba se obtiene cuando la probabilidad de tener la enfermedad antes de la prueba está entre 30 y 70%, pues cuando es menor o mayor la probabilidad, el resultado después de la prueba no va a cambiar sustancialmente la probabilidad de tener o no tener la enfermedad, además serán aplicables a cada individuo en su caso en particular. Por ejemplo, en un individuo hacinado, desnutrido, con pérdida de peso, sudoración nocturna, y tos hemoptoica, la probabilidad de tener tuberculosis sería muy alta antes de investigar BK en esputo y seguiría así a pesar de un resultado negativo. Aquí el examen se haría como requisito, pero muchas veces en regiones endémicas y población de alto riesgo (indígenas) se iniciaría el tratamiento aun sin este resultado. Tendríamos aquí el concepto del va-

## Evaluación de la literatura

lor predictivo positivo dada una prueba negativa el cual estaría influenciado por la prevalencia, con efectos mayores en rangos de 30% a 70% pero menores en prevalencias más bajas o más altas.

Cuando la prueba diagnóstica se utiliza para evaluar el grado de enfermedad más que su ausencia o presencia, la comparación con el patrón puede hacerse más compleja, y se haría mediante la utilización del Kappa. Esta nos muestra la concordancia entre los dos exámenes más allá de lo que podría haber pasado por azar, o por medio de la sensibilidad y especificidad de acuerdo con ciertos valores, en vez de resultados positivos o negativos únicamente. Frecuentemente en observaciones clínicas, imágenes diagnósticas, patología y electrocardiogramas, los resultados pueden cambiar debido a variabilidad entre los observadores o intraobservador; aquí también se podría utilizar el Kappa como estimador de la concordancia, más allá del azar en variables discretas.

Ejemplo: Lozano y colaboradores estudiaron la concordancia entre observadores para establecer la edad gestacional del recién nacido, mediante la prueba de Ballard; compararon las observaciones entre dos pediatras y dos médicos residentes de pediatría de primer año, en 49 recién nacidos. Encontraron un Kappa ponderado de 0.45 y consideran los autores que no es suficiente grado de concordancia más allá del azar, ya que debería haber sido de 0.50 a 0.60 como mínimo para un grado aceptable de concordancia (8).

Así, la eficacia de una prueba estará determinada por la vali-

dez y la concordancia de la prueba, ya que podríamos tener una prueba muy válida (cerca a la verdad) pero muy poco precisa (poca concordancia) o muy precisa (su resultado se repite consistentemente), pero poco válida (lejana de la verdad).

Continuaremos ahora examinando los pasos en la evaluación del artículo de examen diagnóstico:

2. ¿Se evaluó la prueba diagnóstica en una muestra de pacientes que incluyó un espectro apropiado de casos con enfermedad leve a grave, tratada, y no tratada además de individuos con alteraciones diferentes pero que se confunden con frecuencia? Ejemplo, pacientes con artritis reumatoidea con formas leves podrían no quedar incluidos, determinación TSH en pacientes ansiosos, temblorosos o con hipertiroidismo.

3. ¿Se describió adecuadamente la selección de sujetos para esta evaluación así como el filtro que pasaron los pacientes estudiados? Está en relación el VPP con la prevalencia de la alteración blanco, por eso se debe describir adecuadamente el sitio del estudio (hospital de referencia) y el filtro para la selección de las pacientes en los cuales se aplicó la prueba para permitirle al lector decidir si son aplicables a sus pacientes. También se describirán los sujetos controles que no tienen la alteración blanco y que sirvieron para comparación para establecer que las diferencias se deben a la enfermedad en sí y no a otro mecanismo como sexo, edad, raza, dieta u otras variables de confusión.

4. ¿Se determinó la reproducibilidad del resultado de la prueba (precisión) y de su interpretación (variación del observador)? La validez de una prueba deman-

da que se acerque a la verdad (ausencia de sesgo) y que sea precisa (la misma prueba aplicada en el mismo paciente sin que haya sufrido ningún cambio debe dar los mismos resultados), esto es más cierto cuando se requiere experiencia para una mejor realización de la prueba o para su interpretación.

5. ¿Se definió correctamente el término normal para la aplicación de la prueba? Al asignar normal a un resultado los autores deben explicar a qué se refieren y usted debe cerciorarse de que dicha definición es clínicamente sensata, dadas las diferentes definiciones de normalidad utilizadas en medicina (gausiana percentilo, factor de riesgo, deseable culturalmente, diagnóstica), que identifica un rango de resultados más allá de los cuales es posible que una alteración blanco específica esté presente depende de los rangos y límites de corte o terapéutica en la cual se ha visto que los tratamientos son más beneficiosos que dañinos).

6. ¿Se propone la prueba dentro de un conjunto o secuencia de pruebas y se determinó contribución a la validez del conjunto o secuencia? Un ejemplo sería el uso de colposcopia, citología y biopsia en patología del cervix uterino, donde se debe especificar si es todo el proceso o un paso de la secuencia el que está bajo estudio.

7. ¿Las técnicas de la prueba se describieron suficientemente para permitir su exacta repetición? Ejemplos de tales técnicas muchas veces no especificadas incluyen dietas especiales, ingesta de líquidos, drogas que deben evitarse, medición de dolor, transporte y almacenamiento.

8. ¿Se determinó la utilidad de la prueba? ¿El paciente se beneficiará con la prueba? Una prueba útil aportaría tres cosas: un diagnóstico exacto, sustenta la aplicación de un tratamiento específico y eficaz, y por último proporciona un mejor resultado clínico. Este último aspecto es el más difícil de evaluar, pero se podría por medio de estudios aleatorizados, en los que se hace a uno de los grupos la prueba diagnóstica para evaluar si la prueba modificó los resultados clínicos.

#### **LECTURA CRÍTICA DE LA LITERATURA SOBRE TRATAMIENTOS**

Dentro del inmenso caudal de literatura médica disponible hoy en día, una importante proporción se dedica a informar sobre los resultados de nuevas alternativas terapéuticas, descripción de nuevos medicamentos, comparación de tratamientos, demostración de eficacia o efectividad y otros aspectos relacionados con el sano y loable interés de encontrar cada vez respuestas más adecuadas a los innumerables problemas que plantea el cuidado de la salud y la enfermedad de los individuos en una comunidad.

No ha de extrañar por lo tanto que sea éste uno de los aspectos de la literatura médica más apetecido por parte de los lectores y aquél que recibe una mayor relevancia dentro del diseño editorial de las publicaciones periódicas de gran circulación. Como consecuencia de lo anterior, el médico clínico se ve expuesto a un formidable cúmulo de información que debe enfrentar en forma adecuada si pretende obtener algún beneficio real y de utilidad práctica con un buen

aprovechamiento de su tiempo. En los párrafos siguientes se presenta una guía para hacer de este esfuerzo un agradable ejercicio de análisis, con economía de tiempo y rentables beneficios.

Cuando se requiere información sobre un tratamiento, puede el médico apelar a una de varias fuentes: puede consultar sus viejos libros de texto de terapéutica, los que leerá una vez más con deleite y añoranza quizá, sin encontrar la respuesta a su pregunta o sin aprender nada nuevo; puede sonreír al descubrir cómo aquellas recomendaciones allí consignadas parecen desueltas o francamente contraindicadas o simplemente puede recordar algo ya olvidado. Tras su lectura, concluirá que sus libros, un tesoro siempre inapreciable, no le brindan una fuente actualizada y totalmente confiable para enfrentar el problema que le plantea su paciente y se hará el propósito de comprar la última edición disponible en la primera oportunidad que se presente.

Una alternativa casi siempre disponible es consultar a un experto en la materia, quien podrá ofrecer una respuesta rápida, breve y "confiable". En esta forma habremos resuelto el problema en apariencia, pero quizá no habremos aprendido realmente nada al transferir el raciocinio a una segunda persona. Estaremos satisfechos hasta cuando nos enteremos de que nuestro experto consultado lee los mismos libros de texto que elegantemente adornan nuestra biblioteca y que por respeto o prudencia olvidamos preguntarle cuáles, además de su experiencia, eran sus fuentes.

Otra posibilidad se encuentra en la inmensa cantidad de información, que solicita ofrece la in-

dustria farmacéutica y allí se encuentra una amplia gama de colores, formas, variedades y tamaños, que en palabras de sus representantes siempre ofrecen una respuesta aparentemente adecuada. No debe subvalorarse el impacto y la importancia de este método de aprendizaje de nuevas estrategias terapéuticas, como tampoco deben dejar de extremarse los cuidados en su análisis riguroso. Me refiero por supuesto a la estrategia publicitaria de "posicionamiento" de los medicamentos, que utiliza frecuentemente conclusiones de estadíos abiertos u observacionales, cifras impactantes y sentencias concluyentes, que frecuentemente no se apoyan en los estudios que acompañan la publicidad del producto, pero que desde el punto de vista del mercadeo cumplen cabalmente su propósito.

Una estrategia siempre válida de obtener información sobre avances terapéuticos es la asistencia y participación en los cursos, seminarios o congresos de educación continuada, que discuten estos aspectos a condición que ofrezcan evidencia sólida de sus afirmaciones y no simplemente una excursión por los métodos ya citados (asistir a un congreso no es garantía de aprender o confrontar; tan sólo garantiza un certificado).

Finalmente queda el camino de la búsqueda y la lectura crítica de la literatura. Puede parecer el camino más complicado y dispendioso, pero como se verá, su estrategia es infinitamente cercana al método clínico tan caro a nuestros afectos y que realizamos a diario con deleite y eficiencia.

Un ensayo clínico es el método más confiable para probar la efi-



encia de las intervenciones en el cuidado de la salud. Su diseño ha sido concebido para comparar los resultados de introducir una intervención (un nuevo medicamento, un tratamiento no farmacológico, un programa de educación o modificación de hábitos o actitudes) en un desenlace determinado, como sería obtener una mejoría, evitar la recurrencia o las complicaciones, prolongar la vida, maximizar la función, disminuir la discapacidad u otros (9, 10).

De manera sencilla, diremos que en el ensayo clínico, aquellos sujetos que han sido seleccionados como una muestra representativa de la población de interés, una vez satisfacen los criterios de inclusión y exclusión en el estudio y aceptan participar en él tras ser adecuadamente informados sobre los objetivos, riesgos y beneficios, son asignados aleatoriamente, bien a un grupo donde recibirán la intervención bajo experimentación, bien a un grupo control, donde no recibirán tal intervención, para ser seguidos en el tiempo y registrar los resultados, comparando los mismos en ambos grupos (11). Ahora bien, ¿por qué son los ensayos clínicos importantes o necesarios? La literatura abunda en informes sobre tratamientos que gozaron en su oportunidad de aceptación y hoy en día han sido abandonados. Como lo señala McKhann (12), algunos de los pacientes que recibieron estos tratamientos, felizmente no tuvieron consecuencias desafortunadas con su empleo, mientras en otros menos afortunados, hubo empeoramiento de la enfermedad o muerte, pero en todos los casos, decepción y frustración. Atkins (13) ha señalado que si

bien es cierto que algunas intervenciones terapéuticas han producido resultados tan extraordinarios que no han requerido su demostración a través de ensayos clínicos, estos casos son más la excepción que la regla. La mayoría de las nuevas propuestas terapéuticas producen resultados menos espectaculares o márgenes de mejoría más estrechos sobre los tratamientos existentes, por lo que debemos asegurarnos que el nuevo tratamiento es realmente eficaz y su resultado no es simplemente efecto del azar, de diferencias básicas entre los grupos que se comparan, o de una atención, seguimiento y control especiales a los sujetos que lo recibieron. Significa lo anterior que en nuestro interés por obtener información de primera calidad sobre la bondad o utilidad de un tratamiento para una enfermedad dada, nuestra búsqueda deberá enfocarse hacia los experimentos o ensayos clínicos aleatorizados controlados doble ciego (11). Definidos los ensayos clínicos, como la mejor fuente de información, nos detendremos en la estrategia para su revisión: Hace ya dos décadas, el Dr. David Sackett (2, 4) desarrolló en la Universidad de McMaster unas guías para el clínico, con el propósito de facilitar la lectura de estos estudios, formularse un juicio sobre los mismos y optimizar el tiempo, siempre esquivo para el médico que atiende pacientes. Estas guías o pasos hacen referencia por una parte a la validez interna del estudio y por otra a la aplicación de los resultados obtenidos. Presentadas como una serie sucesiva de preguntas son en resumen:

1. ¿Fue la asignación de los pacientes a los tratamientos real-

mente aleatoria?

2. ¿Se informó sobre todos los resultados clínicamente importantes?

3. ¿Eran los pacientes del estudio similares a los nuestros?

4. ¿Se tuvo en cuenta el significado clínico y estadístico?

5. ¿Es la intervención terapéutica factible en la práctica?

6. ¿Fueron tenidos en cuenta para el análisis todos los pacientes admitidos?

Al enfrentar al lector un informe relativo a un ensayo clínico, debe tener en mente estas preguntas y buscar sus respuestas dentro de la estructura del informe bien sea en la sección de material y métodos, bien en los resultados o en la discusión y conclusiones. La asignación aleatoria garantiza que cada paciente que participa en el estudio, tenga la misma probabilidad conocida de ser asignado bien al grupo de tratamiento experimental, bien al grupo control (15). De esta manera se eliminan muchos de los sesgos en que se incurre al no aleatorizar. Aplicando esta guía en forma estricta, el clínico aumenta la eficacia con la cual lee la publicación de su interés, puesto que podrá fácilmente rechazar una buena parte de material, cuando este requisito no se cumple (2, 14).

Frecuentemente los resultados de los experimentos clínicos se consideran satisfactorios, cuando se demuestra una mejoría en el índice o el parámetro que se escogió para comparación, por ser considerado el mejor o el que más fielmente representa el desenlace a medir. Sin embargo, deben examinarse los acompañantes de la mejoría en dicho índice: cuáles fueron los efectos secundarios o colaterales, cuál

la tolerancia o la toxicidad y si se presentaron otros desenlaces clínicamente importantes no previstos inicialmente que pueden modificar la interpretación de los resultados y finalmente cuáles los costos comparativos. En la literatura médica abundan los ejemplos de informes de tratamientos efectivos que pasan por alto la necesidad de incluir todos los desenlaces clínicamente importantes. El lector debe tener siempre en cuenta esta guía, y ser cauteloso antes de aceptar como útil y bondadoso un tratamiento cuando esta información no le ha sido suministrada.

Los autores del estudio deben proporcionar suficiente información sobre las características de los pacientes incluidos: no sólo los aspectos socio-demográficos y clínicos, sino las poblaciones de las cuales derivaron. Tan sólo en esta forma podrá el lector comparar sus propios pacientes con aquellos en los cuales el tratamiento demostró indudable beneficio y apropiar para ellos los resultados. Un punto a tener en consideración es que con frecuencia trasladamos las indicaciones y los resultados de tratamientos a pacientes con características diferentes, con resultados frecuentemente distintos.

Un resultado estadísticamente significativo no es necesariamente clínicamente significativo y viceversa. La significancia estadística nos informa sobre si una diferencia encontrada es real o tan sólo debida al azar. No nos dice si es importante o suficientemente grande.

La significancia clínica hace referencia a la importancia de la diferencia de los resultados en los pacientes tratados y los controles, cuando ella se traduce en

cambios de actitud clínica. Por ejemplo, si los resultados de un estudio nos inclinan a preferir uno sobre otro tratamiento, este resultado es clínicamente significativo. La significancia estadística, frecuentemente informada a través de los resultados de "P" está determinada en gran medida por el número de sujetos que participan en el estudio, por la precisión en la medición y la variabilidad de los datos. El lector debe identificar estos elementos para interpretar los análisis estadísticos y decidir por sí mismo sobre el significado integral de los resultados.

La guía siguiente se refiere a la aplicabilidad de los resultados de los estudios revisados a la propia práctica. Ello requiere en primer lugar, que la intervención o tratamiento de interés haya sido descrito en detalle, para poder ser adecuadamente reproducido. Adicionalmente el tratamiento debe estar disponible y su administración debe seguir principios biológicos aceptados. Este aspecto debe ser cuidadosamente analizado cuando el médico decide "importar" un tratamiento no disponible. Debe asegurarse de poder reproducir cabalmente la intervención.

Finalmente, el lector debe asegurarse de que el trabajo incluya información sobre todos los pacientes admitidos al estudio, independientemente del número final sobre el cual se basan los resultados y los análisis. Esta información incluye los pacientes que no terminaron el estudio, porque se retiraron, por efectos de intolerancia o toxicidad, por pérdidas en el seguimiento, por complicaciones o por muerte, entre otros. Una estrategia rápida y útil es comparar el número

de pacientes que ingresan al estudio y el número final sobre el cual se realiza el análisis de los resultados. Si estos números no coinciden, el informe debe explicar suficientemente estas diferencias. Un informe que base sus resultados en una proporción pequeña de pacientes bajo estudio debe ser interpretado con especial cuidado y cierto escepticismo.

Colocando juntos los pasos anteriores el lector deberá asegurarse de que los informes sobre los ensayos clínicos aleatorizados le proporcionen información adecuada acerca del diseño, el desarrollo, el análisis y la interpretación de los resultados (16). Si estas condiciones se satisfacen, el tiempo empleado en la lectura será recompensado, independientemente de los resultados que el estudio ofrezca.

Estas guías a pesar de haber sido diseñadas hace cerca de veinte años, y de ofrecer una orientación tanto al lector como al autor que quiere ser leído, sus principios no son suficientemente tenidos en cuenta en un gran número de las publicaciones disponibles hoy en día. Recientemente, los comités editoriales de las más prestigiosas publicaciones han recomendado incluir en el resumen una información más precisa acerca de los aspectos de diseño, conducción, análisis e interpretación de los resultados de los ensayos clínicos. Estos resúmenes incluyen subtítulos que introducen esta información.

Un comentario adicional merece la referencia a la discusión sobre el tamaño de la muestra del estudio (17). El tamaño de la muestra es en gran parte responsable del poder del estudio, esto es, de la

## Evaluación de la literatura

capacidad de demostrar diferencias significativas cuando realmente existen y es matemáticamente el complemento del error tipo II o tipo beta, que los autores frecuentemente omiten reportar. Este punto ha sido motivo de análisis recientemente y ha permitido demostrar cómo un buen número de los estudios adelantados que reportan resultados negativos (o ausencia de diferencia en tratamientos), se deben al hecho de no tener poder suficientemente para demostrarla (9). De allí la recomendación de preferir, cuando sea posible o factible, estudios con un tamaño de muestra suficiente.

Al revisar la sección de resultados, el lector debe inicialmente identificar aquel desenlace que se pretendía medir como el más importante y verificar que los resultados apunten hacia este desenlace en particular, sin perjuicio de que los autores aporten información sobre otros resultados secundarios.

En el desarrollo de un estudio clínico controlado, es posible que no todos los pacientes asignados aleatoriamente a un grupo permanezcan dentro del mismo. Pueden cambiar de grupo, violando el protocolo o por sesgo de contaminación. El análisis debe tener en cuenta preferencialmente la asignación aleatoria que los pacientes recibieron al inicio del experimento (principio de "intención de tratar"). Cuando el análisis se realiza sobre la intervención finalmente recibida, las conclusiones pueden resultar erróneas. Los autores pueden, sin embargo, estar más interesados en la eficacia del tratamiento. En este caso, deben informarlo y justificarlo suficientemente (16). Ello le per-

mitirá al lector conservar su distancia frente al análisis y confrontar su propia expectativa sobre los resultados

Actualmente la mayoría de los experimentos clínicos incluyen la colaboración de un experto o especialista en bioestadística. La ventaja de esta participación se incrementa si se cuenta con su concurso desde el diseño mismo del estudio y no tan sólo en la parte final del análisis de los datos. Ello dará más claridad sobre el tipo de variables por medir, las escalas por utilizar y los procedimientos por emplear. Si bien su participación no garantiza resultados libres de sesgos, indudablemente su contribución a la fidelidad de los datos y a su correcto tratamiento, facilitará al lector la comprensión de los resultados. Los procedimientos estadísticos del análisis deben ser completamente enunciados y referenciados. El empleo de unas técnicas sobre otras debe estar justificado y se debe proporcionar la información suficiente para que el lector interesado pueda por sí mismo realizar los análisis comparando sus resultados con los informados.

El valor mágico de la "P" y su calificación como significativa o no significativa, debe ser reemplazado por su valor absoluto y complementado con la medición de los intervalos de confianza cuando sea posible, ya que estos últimos, al acompañar al estimador, brindan una información más completa y útil y se acercan más a la significancia clínica (19).

Cuando un experimento clínico adecuadamente conducido presenta resultados referidos al desenlace previamente identificado como el más importante, con un

correcto tratamiento analítico y una interpretación apropiada, el lector puede darse por bien servido y sentirse satisfecho, pero ha de tener especial cuidado, si los autores derivan de sus resultados conclusiones sobre problemas no planteados al inicio del estudio, o si se adentran en comparaciones no planeadas.

Si bien estos puntos parecen imponer unos requisitos excesivos a los ensayos clínicos controlados, constituyen una guía para autores, editores y lectores, con el propósito de hacer más uniforme su presentación, subrayando la validez interna de los estudios y aumentando la confianza de los lectores. Estos a su vez, podrán acometer la lectura de una manera más crítica, desechando aquellos trabajos que se aparten del informe estructurado u omiten información acerca del diseño, desarrollo, análisis e interpretación de los resultados. Esta propuesta tiene la ventaja de convertir al lector en un verdadero interlocutor, con un papel más crítico y más activo, con mayor capacidad de juzgar, comparar, rebatir o aun rechazar la evidencia presentada. Progresivamente los lectores aumentarán sus exigencias y tanto autores como editores deberán estar más atentos a la calidad de sus informes, al tipo de información que brindan y dispuestos a estar expuestos a la lectura crítica de sus materiales. Estas notas no se han ocupado de los aspectos éticos de los estudios, de la elección de variedades de diseño, de lo relevante de las preguntas y se ha detenido en subrayar aquellos aspectos que deben guiar la lectura crítica de la literatura médica relativa a la eficacia de los tratamientos que

es a diario presentada en el creciente mundo de las publicaciones médicas, que llenan nuestras bibliotecas y compiten por nuestro escaso tiempo.

Sin embargo, de manera similar a lo que hacemos al construir la historia clínica de la enfermedad en cada uno de nuestros pacientes, la lectura crítica debe empezar por la identificación del problema que se pretende estudiar, que equivale al motivo de consulta, a los conocimientos o experiencias previas sobre el tema, con sus aspectos definidos y aquellos no aclarados, que equivaldría a los antecedentes, a la definición de las variables, los desenlaces, la forma adecuada de medirlos o compararlos que sería similar a la historia de la enfermedad actual, a los métodos empleados, que al igual que la revisión por sistemas, examina en varios campos la aproximación a la respuesta que se busca, a la conducción misma del experimento, que requiere el rigor del examen minucioso, registrando tanto los hallazgos positivos como los negativos, con un resumen de los mismos, que al igual que los resultados, permitan acercarse a un diagnóstico o a unas conclusiones, quizá dejando un espacio para explicaciones alternativas o diagnósticos diferenciales o abriendo el compás para investigaciones

adicionales o lecturas complementarias.

La lectura crítica de la literatura relativa a la eficacia de tratamientos así concebida se constituye en un reto clínico, en un ejercicio de la inteligencia, en un esfuerzo diagnóstico, y por ello mismo, lleno de enseñanzas útiles y constructivas. El conocimiento así adquirido nos dará evidencias más sólidas, mayor peso a nuestras afirmaciones y una prudente confianza en nuestras futuras decisiones.

Es deseable que en todos aquellos escenarios donde se realicen clubes de revistas, se ejercite esta estrategia y se experimenten sus beneficios y que los docentes se ejerciten en su práctica en beneficio de sus pacientes y sus estudiantes.

#### REFERENCIAS

1. **Guyatt C, Sackett D, Cook D.** Users Guide to the Medical Literature. *JAMA* 1993; **270**: 2598-2601.
2. **Sackett DL, Haynes RB, Guyatt GH, P. Tugwell.** *Clinical Epidemiology*. 2nd ed. Boston Mass: Little Brown Co; 1991.
3. **Fletcher RH, Fletcher SW, Wagner EH.** *Epidemiología Clínica*. Barcelona, Ediciones Consulta 1989 2a. edición.
4. **Moreno L, Cano F, Romero H.** *Epidemiología Clínica*. México. Nueva Editorial Interamericana 1994 2a. edición.
5. **Ransohoff DF, Feinstein AR.** Problems of Spectrum and Bias in Evaluating the Efficacy of Diagnostic Tests. *N Engl J Med*. 1978; **299**: 926-930.
6. **Goncalves LF, Jeantv P, Piper JM.** The Accuracy of Prenatal Ultrasonography in Detecting Congenital Anomalies. *Am J Obstet Gynecol* 1994; **171**: 1606-1612.
7. **Gaitán H, Castañeda E, Medina M, Vargas CI, Angel E.** Diagnóstico de la Enfermedad Pélvica inflamatoria y sus complicaciones en el Instituto Materno Infantil de Santafé de Bogotá en 1992-1993. Manuscrito en preparación.
8. **Lozano JM, Martínez A, Pardo R, Vera JF.** Concordancia entre observadores en la determinación de la edad gestacional por medio de la prueba de la Ballard. *Actual Pediatr FSFB* 1991; **11**:6-10.
9. **Byar DP, Simon RM, Friedewal WT, et al.** Randomized clinical trials. *N Engl J Med* 1976; **295**: 74-80.
10. **Mosteller F.** Innovation and evaluation Science 1981; **221**: 881-885.
11. **Kramer MS.** *Clinical trials*. En Kramer MS *Clinical Epidemiology and biostatistics* Springer-Verlag Berlin Chapter 1988; **7**: 78-92
12. **McKhann GM.** The trials of clinical trials. *Arch Neurol* 1989; **46**: 611-614.
13. **Atkins H.** Conduct of a controlled clinical trial. *Br Med J*, 1966; **2**: 377-384.
14. **Sackett DL.** How to read clinical journals. *CMA Journal* 1981; **124**: 1156-1162.
15. **Zelen M.** The randomization and stratification of patients to clinical trials. *J. Chron Dis* 1974; **27**: 365-375
16. The standards of reporting trials group. A proposal for structured reporting of randomized controlled trials. *JAMA* 1994; **272**: 1926-31.
17. **Dennis RJ.** Cómo estimar el tamaño de muestra necesaria en investigaciones humanas. *Acta Med Colomb* 1989; **14**:92-99.
18. **Freiman JA, Chalmers TC, Smith H, Kuebler RR.** The importance of beta, the type II error, and sample size in the design and interpretation of the randomized control trials survey of 71 negative trials. *N Engl J Med* 1978; **299**: 690-694
19. **Gardner MJ, Altman DG.** Estimating with confidence in Statistics with Confidence, published by the British Medical Journal London 1989 Chapter 1.