

Reumatología Basada en Evidencias.

Evaluación Crítica de los Ensayos Clínicos

Pablo Riedemann (M.Sc.)

Unidad de Reumatología y CIGES*, Facultad de Medicina,
Universidad de la Frontera, Temuco, Chile

Introducción

Los Ensayos Clínicos corresponden al diseño de investigación más apropiado para evaluar la eficacia y efectividad de distintas medidas terapéuticas y/o preventivas. En este artículo queremos poner en una perspectiva práctica los aspectos metodológicos de este tipo de diseños y a partir de ello poder hacer más provechosa la lectura de esta clase de artículos médicos.

Partamos simplemente por preguntarnos: **¿por qué leemos artículos médicos?**. Sin duda las motivaciones serán diferentes e incluirán, entre otras, la curiosidad, la necesidad de información para realizar un proyecto de investigación, el tener que preparar una clase para alumnos o una conferencia para colegas. Sin embargo, indudablemente nuestra motivación fundamental y más frecuente es la necesidad de información para resolver problemas de nuestros pacientes.

Satisfacer esta necesidad se ve dificultada por varios hechos. Por un lado existe una avalancha de publicaciones biomédicas, que hace extremadamente difícil, si no imposible, el poder leer todo lo que se publica, aun de un aspecto muy puntual de la medicina. Si tomamos como ejemplo una patología de frecuente consulta, como la Artritis Reumatoide, y nos centramos sólo en su tratamiento, solamente durante el año 2000 en MEDLINE aparecen aproximadamente 1.300 publicaciones. Para revisarlas todas deberíamos leer diariamente

te casi cuatro publicaciones, lo que implica a lo menos dos horas de lectura diaria, sólo para ese tema.

Otro aspecto que dificulta nuestro acceso a la información es la creciente dificultad en comprender la metodología usada en estos estudios, lo que se ve acrecentado por el limitado tiempo que tenemos para dedicar a la lectura. De lo anterior se deduce que necesitamos herramientas y guías que nos ayuden a simplificar esta tarea y hacerla más abordable.

Decisiones clínicas y evidencias

Todos hemos escuchado el aforismo “la medicina ES UNA CIENCIA y un ARTE”; sin embargo, no todos conocen la respuesta que algunos dan a esto, diciendo: “sí, es una ciencia y un arte, pero ES UNA CIENCIA llena de incertidumbres y el ARTE de tomar decisiones sin la adecuada información”.

Podríamos pensar que esto es una exageración, que los médicos somos extremadamente rigurosos en nuestro actuar y siempre decidimos sobre bases sólidas. Sin embargo, y lamentablemente, aun frente a situaciones simples y frecuentes, esto no siempre es así.

Veamos un par de ejemplos para ilustrar este aspecto. Imaginen que son médicos que trabajan en un Servicio de Urgencia (efectivamente, muchos de ustedes lo hacen o lo han hecho), y llega a consultar un hombre de 33 años por un dolor abdominal de reciente comienzo. Al examinarlo, el abdomen esta rígido y sensible a la palpación; les parece que incluso tiene el signo de Blumberg positivo. El interno que los acompaña, al ver tan adolorido al paciente, les propone administrarle un

Manuel Montt 112, Casilla 54-D, Temuco
Fono 45 – 325755. Fax 45 – 325741
E-mail: riedeman@ufro.cl

* CIGES: Centro de Capacitación, Investigación y Gestión para la Salud Basada en Evidencias

analgésico narcótico. **¿Cuál sería la respuesta que le darían?**

Lo más probable es que rechacen la sugerencia e incluso llamen la atención al interno por proponer una terapia que pudiera encubrir un cuadro serio y confundir el diagnóstico (otros pudieran tener reacciones aún más viscerales, por lo que demos gracias de no ser internos).

Esta respuesta parece del todo concordante con lo que se nos enseñó en nuestros años de estudio, así como la práctica habitual de quienes trabajan en urgencias. Pero veamos qué dicen las evidencias al respecto.

Attard y colaboradores (1) estudiaron a 100 pacientes ingresados a un servicio de urgencia con dolor abdominal significativo. Los pacientes fueron distribuidos aleatoriamente a recibir un analgésico narcótico o placebo. Se evaluó el efecto terapéutico en cuanto al alivio del dolor, el confort del paciente, la certeza diagnóstica y la decisión del manejo por parte del cirujano (observar u operar).

El alivio del dolor fue significativamente mejor en los pacientes que recibieron el opiáceo; 96% de los tratados con opiáceo se sentían confortables luego de su administración vs sólo 18% de los que recibieron placebo. Lo que es aún más importante es que el error en la decisión de manejo se produjo sólo en 2 pacientes tratados con opiáceos vs 9 de los pacientes tratados con placebo.

En el año 1996, Pace y col. (2) realizaron un estudio similar en el que de 71 pacientes estudiados tuvieron errores de diagnóstico sólo en 3 pacientes de cada grupo.

Cambemos de patología y piensen ahora que consulta una mujer joven, de 26 años, por dolor en el ojo derecho. Refiere que estaba jugando voleibol y le dieron un pelotazo en el ojo. Al examinarla ustedes constatan que tiene una pequeña erosión corneal. **¿Cómo la tratarían?**

Lo más probable es que le indiquen algún analgésico y oclusión ocular por algunos días (eso es lo que nos enseñaron). Seguramente deben estar pensando que probablemente haya evidencias en contra.... ¡y acertaron! Ya en 1960 fue publicado un trabajo comparativo que evaluaba la evolución de pacientes con erosiones corneales. Cubrir o no el ojo afectado no determinó diferencias en la curación, y el autor proponía que las erosiones simples no requerían cubrir el ojo (3).

Más recientemente, Kaiser (4) realizó un es-

tudio en 201 pacientes con erosiones corneales que fueron aleatorizados a recibir o no oclusión corneal. Concluyó que aquellos que no recibieron parche tuvieron menos dolor y mejoraron más rápido.

Decisiones clínicas y necesidades de información

Si bien éstos son sólo algunos ejemplos, podemos sospechar que nuestro actuar como médicos no siempre está respaldado por evidencias científicas y por tanto es presumible que frecuentemente debiéramos necesitar información para resolver los problemas de nuestros pacientes.

Curiosamente, si nos preguntan al respecto, la percepción es que requerimos de esta información sólo ocasionalmente. Este aspecto fue revisado por Cowell y col. (5) evaluando las necesidades reales de información en la práctica clínica habitual. Estudiaron a 47 médicos internistas, en quienes evaluaron medio día de trabajo habitual. Antes de iniciar el estudio le preguntaron a cada uno: **¿Cuántas veces por semana necesita información para resolver el manejo de un paciente?**; la respuesta fue que la necesitaban en promedio una vez por semana.

En esta experiencia clínica se les pidió a los médicos que realizaran su trabajo habitual de atención ambulatoria y una vez que terminaban de atender a cada paciente se les preguntaba: **¿tiene dudas respecto del paciente que acaba de atender?**. Al final del medio día de trabajo se les preguntaba respecto de las fuentes de información usadas y las dificultades que tuvieron en conseguir esa información.

Este grupo de 47 internistas atendió 409 pacientes en medio día de trabajo, y les surgieron ¡269 dudas!. Dicho en otras palabras, un promedio de 6 dudas por médico o 2 dudas por cada 3 pacientes. Las áreas que más frecuentemente presentaron dudas fueron el diagnóstico y los tratamientos (70% de las preguntas).

Por tanto, es probable que necesitemos información más de lo que creemos (en 2 de cada 3 pacientes), y si vemos 60 enfermos en una semana, en 40 de ellos esta información podría afectar nuestra decisión médica. Cada uno de nosotros puede preguntarse: ¿en cuántos de estos pacientes tenemos el tiempo y los medios para obtener esta información?.

Decisiones terapéuticas - lectura crítica

Centraremos las siguientes líneas en las decisiones terapéuticas. Cuando el diagnóstico está hecho, o nos parece muy probable: **¿debemos iniciar tratamiento?**; y como muchas veces ocurre, si existe más de uno, **¿cuál es el mas apropiado?**.

Lamentablemente, pocas terapias tienen efectos tan notables como la penicilina en la neumonía neumocócica o la cirugía en la apendicitis aguda.

La mayoría de las veces los efectos de los tratamientos son relativamente pequeños. A veces el cambio será una reducción en la mortalidad de 2% o 3%, o una mejoría en la calidad de vida, pero **¿cómo valorar adecuadamente estos resultados y decidir de manera informada?**.

Desde hace varios años se han desarrollado diferentes pautas, herramientas y recursos que permiten ser más eficientes en la lectura de artículos clínicos (6).

A esta estrategia sistemática se la ha llamado apreciación crítica de la literatura médica o simplemente lectura crítica; ella pretende ayudarnos a lidiar con esta avalancha de publicaciones, orientándonos a lo que debemos prestar atención y a tomar la decisión de cambiar nuestra práctica habitual y adoptar la nueva terapia que se nos ofrece cuando ello parece razonable.

La lectura crítica corresponde a un conjunto de herramientas metodológicas y estadísticas básicas, para ser usadas por el médico clínico en la evaluación de las publicaciones médicas. Entre otras, existen pautas para evaluar artículos sobre causalidad, curso clínico y pronóstico, tests diagnósticos, terapia, análisis económico y revisiones de la literatura.

Como mencionáramos, en el caso de las decisiones terapéuticas (y también preventivas), la evidencia de mejor calidad es aquella que proviene de los Ensayos Clínicos. Para leer críticamente una publicación de este tipo, son tres las preguntas fundamentales que debemos plantearnos (7):

- I. **¿Son válidos los resultados del estudio?**
- II. **¿Cuáles fueron los resultados del estudio?**
- III. **Los resultados ¿me serán útiles para el cuidado de mis pacientes?**

Intentaremos ilustrar este aspecto con un ejemplo real y reciente. Casi todos hemos leído el

llamado estudio VIGOR, ensayo clínico comparativo para evaluar la seguridad de Rofecoxib vs Naproxeno en pacientes con artritis reumatoide (8). Antes de continuar es necesario dejar muy en claro que la intención de este artículo no es hacer una valoración o emitir un juicio respecto del estudio, sino que sólo usarlo como ejemplo real. Estamos seguros de que cada uno de ustedes tendrá su propia opinión al respecto y esperamos que estas líneas les ayuden en ello.

En esa investigación se estudió a 8.076 pacientes con artritis reumatoide de 50 años o más (40 años si usaban esteroides), los cuales fueron aleatorizados a recibir ya sea 50 mg de Rofecoxib o 500 mg de Naproxeno dos veces por día. Estos pacientes fueron seguidos por un tiempo promedio de nueve meses, vigilando, entre otros, la aparición de perforaciones, úlceras o sangramientos.

La primera pregunta que nos debemos plantear es: **¿son válidos los resultados del estudio?**. Dicho de otra manera, debemos preguntarnos si es que los resultados que se reportan son verdaderamente representativos de la situación estudiada. Esto se refiere a la credibilidad que tiene el estudio desde el punto de vista metodológico y la rigurosidad que tuvieron los investigadores en dar cumplimiento a algunos aspectos fundamentales de toda investigación terapéutica.

Para ayudarnos a determinar si es que los resultados del estudio son válidos, existen algunos criterios denominados primarios:

- 1) **La asignación de los tratamientos a los pacientes, ¿fue aleatoria?**
- 2) **¿Se tuvo en cuenta a todos los pacientes incluidos en el ensayo y se los ha considerado en la conclusión?**
- 3) **¿Se ha realizado un seguimiento completo?**
- 4) **¿Se analizaron los pacientes en los grupos en que fueron asignados aleatoriamente?**

Veamos cada uno de estos aspectos con un poco más de detención.

1) **¿La asignación fue aleatoria?**: La importancia de esto radica en que los clínicos, así como también los investigadores, al enfrentar a un paciente o realizar un estudio, siempre tenemos una hipótesis respecto de cuáles debieran ser los resultados o cuáles debieran ser los efectos del tratamiento que deseamos evaluar. Si la asignación

de los pacientes a los grupos en estudio, o alternativas terapéuticas en comparación, fuera hecha con el conocimiento del clínico o el investigador, aun sin quererlo y de manera subconsciente, podría tender a favorecer aquella opción terapéutica que desea evaluar; por ejemplo, podría destinar al grupo de la nueva terapia a aquellos pacientes en quienes crea que la posibilidad de éxito es mayor, o que van a cumplir mejor las indicaciones.

Ello introduciría un sesgo potencial que podría hacer variar los resultados. Por lo tanto, idealmente, en todo estudio terapéutico o preventivo, se debería considerar como diseño de investigación el Ensayo Clínico, que por definición incluye la distribución aleatoria de los pacientes a los grupos en estudio (aleatorización). Este procedimiento, cuando es correctamente realizado, permite que todos los individuos que son incorporados al estudio tengan una similar posibilidad de recibir la terapia en estudio o la terapia de comparación. Por otro lado, y cuando el tamaño muestral es suficiente, permite que se obtengan grupos comparables en cuanto a las características de los individuos que los componen.

Las siguientes preguntas están interrelacionadas entre sí y tienen que ver con el apropiado manejo de la información:

- 2) **¿Se tuvo en cuenta a todos los pacientes incluidos en el ensayo y se los ha considerado en la conclusión?**
- 3) **¿Se ha realizado un seguimiento completo?**
- 4) **¿Se analizaron los pacientes en los grupos en que fueron asignados aleatoriamente?**

Si del total de pacientes que fueron incorporados a la investigación algunos de ellos salieran del estudio por diferentes razones (falta de eficacia o efectos adversos, por ejemplo) o algunos se perdieran durante el seguimiento, y al momento del análisis esto no fuera considerado apropiadamente, ello podría sesgar los resultados. Imaginemos que en el grupo de los tratados con una nueva terapia, un número importante abandona el estudio por falta de eficacia y en el análisis sólo se incluye a aquellos que completaron la investigación, este nuevo tratamiento podría parecer mejor de lo que en realidad es.

Si consideramos los aspectos mencionados, el lector podría volver a revisar el estudio VIGOR. Resumidamente, en este estudio, los pacientes incorporados, luego de un periodo de 3 a 14 días

sin AINE, fueron asignados aleatoriamente a recibir Rofecoxib 50 mg una vez por día o Naproxeno 500 mg dos veces por día.

Los pacientes fueron estratificados según la presencia o no de historia de úlceras, sangramientos o perforaciones previas. El estudio fue ciego usando placebo idéntico para cada droga activa. Los pacientes fueron seguidos en promedio durante nueve meses (entre cinco y 13 meses).

Del total de pacientes, un 71,1% tomó sus medicamentos hasta el final del estudio, pero todos los pacientes fueron considerados al momento del análisis.

De acuerdo a lo mencionado, claramente el estudio que estamos usando de ejemplo cumple con los criterios primarios de validez.

Con relación a este aspecto de validez del estudio, existen también unos criterios secundarios:

- 1) **¿El estudio fue ciego?**
- 2) **¿Los grupos eran comparables al inicio del estudio?**
- 3) **¿Aparte de la intervención en estudio, ¿los grupos fueron tratados de la misma forma?**

En cualquier investigación existen múltiples fuentes potenciales de sesgo, entendiéndose como tal cualquier error sistemático que hace que los resultados se aparten de la verdad. Con relación al ciego, si por ejemplo un paciente o su médico conocieran el tratamiento que le ha sido asignado, y éste no correspondiera a sus expectativas (deseaba estar en el grupo de la nueva terapia y le tocó placebo), podrían generarse conductas que alterarían los resultados; por ejemplo, el paciente podría abandonar el estudio, o por su cuenta usar otro medicamento, o el médico le podría prescribir otro medicamento a cambio, etc.

Si por otro lado el paciente hubiera recibido la nueva terapia y lo sabe, esto podría tener un efecto en sugestionar al paciente y predisponerlo a reportar efectos favorables o minimizar los efectos adversos.

Por ello, todo estudio terapéutico debería explicitar si fue o no ciego, y si no lo fue, dar razones de peso para hacerlo aceptable.

2) **¿Los grupos eran comparables al inicio del estudio?:** La existencia de desbalances en las características de los grupos en comparación podría determinar diferencias en los resultados, las cuales no fueran debidas al efecto de las alternati-

vas terapéuticas en comparación, sino más bien a las características de los individuos.

Como se mencionara, en general, y en estudios de tamaño muestral suficiente, un proceso de aleatorización bien realizado tenderá a que se obtengan grupos comparables, pero siempre es deseable constatar que ello ha ocurrido. En el estudio que estamos usando de ejemplo, en la Tabla 2 de la publicación original se presentan las características de los pacientes al comienzo del estudio. Considerando aspectos como edad, sexo, raza, duración de la enfermedad, % con FR (+), uso previo de ciertos medicamentos, etc., los pacientes eran en todo comparables.

Si en un estudio existieran diferencias importantes en ciertas características iniciales de los sujetos y esto no fuera considerado al momento del análisis, obviamente ello podría determinar diferencias que en realidad son debidas a esos desbalances y no a efectos producidos por el tratamiento que se evalúa.

3) Aparte de la intervención en estudio, **¿los grupos fueron tratados de la misma forma?**: Este aspecto es otra fuente potencial de sesgo. Si un tratamiento para el dolor y la inflamación fuera inferior a otro, pero los pacientes pudieran automedicarse según necesidad y ello no fuera registrado, al valorar el resultado respecto del alivio del dolor podría aparecer que los individuos de ambos grupos tuvieron un resultado similar. Sin embargo, la realidad es que el alivio del dolor en el grupo que recibía el medicamento de inferior efecto fue producto de la automedicación y no efecto de la terapia original. Este fenómeno es conocido como cointervención, y el lector debería considerar este aspecto al evaluar los resultados de un ensayo clínico.

Si los aspectos anteriores se cumplen, especialmente los criterios primarios, podremos considerar que los resultados del estudio son válidos, y por tanto puede ser útil el que sigamos leyendo. Si, por otro lado, no se cumplieran estos criterios y entonces el estudio no tiene resultados válidos, es mejor invertir nuestro tiempo en jugar un rato con los hijos, sobrinos o nietos y no perder tiempo en un artículo no válido.

Si consideramos que los resultados del estudio que estamos leyendo pueden ser considerados válidos, entonces viene la siguiente pregunta:

II) ¿Cuáles fueron los resultados del estudio?

En relación con esta interrogante, dos son los aspectos fundamentales:

- 1) **¿Cuál fue la magnitud del efecto del tratamiento?**, y
- 2) **¿Cuál ha sido la exactitud en la estimación del efecto del tratamiento?**.

Necesitaremos ahora ayudarnos de algunas matemáticas básicas. Toda investigación realiza una estimación (medición) referente a la ocurrencia de un fenómeno: número de muertes, frecuencia de efectos adversos, magnitud de un cierto beneficio (alivio del dolor, por ejemplo), etc. Esto se logra estudiando a un grupo o muestra de la población, y por lo tanto los resultados representan una estimación de la realidad en esa población.

Lo primero que interesa conocer es la magnitud o tamaño del efecto; dicho en otras palabras, qué tanto mejor o más efectivo es un determinado tratamiento sobre otros.

Usemos como ejemplo el efecto que puede tener un fármaco sobre la mortalidad, e imagine-mos que en el grupo control muere un 20% (0,20) y que en el grupo tratado fallece el 15% (0,15). Interesará entonces hacer una evaluación del riesgo de morir. Existen varias formas de valorar esto:

- a) Reducción Absoluta de Riesgo (**RAR**)
- b) Riesgo Relativo (**RR**)
- c) Reducción Relativa de Riesgo (**RRR**)

La **RAR** es el valor absoluto de la diferencia aritmética entre las tasas de ocurrencia del evento (muerte) en los controles comparados con los tratados. Los pacientes en el grupo control tienen un riesgo de morir de 20%, y los tratados un riesgo de 15%. Por tanto, la $RAR = (20 - 15) = 5\%$.

El **RR** es el riesgo de eventos (morir) para los pacientes que reciben el tratamiento relativo al riesgo de los controles. En este caso el riesgo con tratamiento es de 15% (0,15) y de los controles 20% (0,20). Entonces el **RR** será igual a $(0,15/0,20) = 0,75$. Dicho en otras palabras, si el riesgo de morir para los pacientes en el grupo control es de 1, para los tratados el riesgo se reduce a 0,75.

Finalmente, la **RRR** corresponde a la reduc-

ción porcentual de la tasa de ocurrencia de un evento desfavorable (muerte) en el grupo tratado (o grupo experimental) en relación con el grupo control. En nuestro ejemplo la Tasa de Eventos (muertes) en Tratados (TET) = 15% y en los controles (TEC) = 20%. Entonces la RRR = $(TET - TEC) / TEC = (15 - 20) / 20 * 100 = - 25\%$. Dicho en palabras, la RRR de morir de los tratados con respecto de los controles es de 25%.

Veamos ahora qué pasaría si en vez de riesgos de 15% y 20%, los riesgos de morir fueran de 1,5% para los tratados y de 2% para los controles. Ustedes pueden practicar realizando los cálculos de la manera enunciada, pero para simplificar la tarea ellos se presentan en la Tabla 1.

De la tabla es interesante observar que si bien la RAR en el segundo ejemplo (0,5%) es 10 veces menor que en el ejemplo original, tanto el RR como la RRR son exactamente iguales. El mensaje aquí es que debemos saber leer entre líneas y comprender que si bien el RR y una RRR son importantes, también puede serlo la RAR, especialmente cuando comparamos opciones terapéuticas.

Volvamos a nuestro estudio de ejemplo. De los objetivos, uno de los más importantes era la reducción en la tasa de eventos gastrointestinales (úlceras, sangramientos y perforaciones). Después del seguimiento promedio de nueve meses, se reportaron 2,1 eventos confirmados por 100 pacientes/año en los tratados con Rofecoxib comparados con 4,5 eventos por 100 pacientes/año en los tratados con Naproxeno. Si expresamos estas cifras siguiendo el modelo que hemos presentado, el riesgo en los tratados (Rofecoxib) es de 2,1 (0,021) y de los controles (Naproxeno) 4,5 (0,045). En la Tabla 2 se presentan los cálculos de riesgo para esta situación.

Si alguien revisa el estudio original, se reporta un RR de 0,5 que corresponde a una aproximación del cálculo exacto y probablemente resulta de más fácil comprensión. Expresado en palabras, los pacientes tratados con Rofecoxib tienen una RAR de sufrir eventos gastrointestinales de 2,4%, un RR de 0,5 al compararlos con los tratados con Naproxeno, y respecto de estos últimos, tienen una RRR de 53%.

Tabla 1			
Estudio	RAR	RR	RRR
Control 20% (0,20)	20 - 15 = 5%	0,75	25%
Tratados 15% (0,15)			
Control 2% (0,02)	2 - 1,5 = 0,5%	0,75	25%
Tratados 1,5% (0,015)			

Tabla 2				
Riesgo Control Naproxeno (X)	Riesgo Tratados Rofecoxib (Y)	RAR X - Y	RR Y / X	RRR (TET-TEC)/TEC
4,5 %	2,1 %	2,4 %		
0,045	0,021	0,024	0,47	53%

En relación con los resultados, además de la estimación del tamaño del efecto, nos debiera interesar también cuál ha sido la exactitud en la estimación del efecto del tratamiento (9).

El verdadero tamaño del efecto, o en otras palabras, la verdadera reducción del riesgo, nunca se sabrá; lo mejor que podemos tener es la estimación que nos puedan entregar ensayos clínicos muy bien realizados. Sin embargo, en toda medición hay cierto grado de incertidumbre y a eso se refiere el tratar de conocer la exactitud en la estimación del tamaño efecto.

Todos estamos familiarizados con el valor de p , y muchas veces nos dejamos impresionar cuando su valor es extremadamente pequeño ($p < 0,001$, por ejemplo). Pero dicho en términos sencillos, ¿que es el valor de p ?. Este número, muchas veces casi mágico, sólo representa una cuantificación del grado de incertidumbre que tiene un resultado. Por ejemplo, decir que el valor de p es $< 0,05$, significa que la probabilidad de que el resultado observado sea producto del azar y no corresponda a una diferencia real es de menos de 5%.

En el estudio que hemos usado como ejemplo, para el RR observado se reporta un valor de $p < 0,001$. Esto quiere decir que la probabilidad de que ese resultado sea sólo producto del azar (y no una diferencia real) es menor al 1%.

Nos damos cuenta, entonces, de que en relación con la exactitud en la estimación del tamaño del efecto, mientras más pequeño sea el valor de p mayor será la exactitud de la estimación o resultado que observamos.

Otra manera de cuantificar esta exactitud es el llamado intervalo de confianza (IC), habitualmente expresado como el IC 95%. Al resultado obtenido se le conoce también como estimación puntual, dando a entender con ello que el verdadero valor de la reducción del riesgo está alrededor de esta estimación puntual.

Una manera de cuantificar el margen en el cual se encuentra este valor real es el llamado IC del 95%. Esto puede entenderse como los valores entre los cuales se encuentra el verdadero valor del RR el 95% de las veces. En el caso del estudio que hemos venido usando como ejemplo, se reporta un RR de 0,5 con un IC 95% de 0,3 a 0,6; esto quiere decir que la verdadera reducción de riesgo estará entre 0,3 y 0,6 en el 95% de las veces.

Respecto de la exactitud o certeza en la determinación puntual, mientras más pequeño es el margen entre los valores del IC, mayor será la exactitud.

Consideremos ahora la posibilidad de revisar otro estudio donde se reporta una estimación puntual de RR = 0,8, pero el intervalo de confianza sea entre 0,3 y 1,5, además de considerar el amplio margen que incluyen estos valores, ellos incluyen 1, lo cual significa que puede no haber ninguna reducción de riesgo, e incluso haber un mayor riesgo.

Las preguntas finales que siempre debemos hacernos cuando revisamos un estudio sobre tratamiento o prevención son referentes a la aplicación de sus resultados. Tres aspectos son importantes:

- 1) **¿Los resultados son aplicables a mis pacientes?**
- 2) **¿Se consideraron todos los resultados clínicamente importantes?**
- 3) **¿Los beneficios potenciales superan los riesgos (efectos colaterales) y costos?**

El primer aspecto tiene que ver con qué tan similares son sus pacientes habituales comparados con los que participaron en el estudio. Si de acuerdo a los criterios de selección reportados sus pacientes podrían haber sido uno más de los incorporados al estudio, no hay duda de que puede aplicar a ellos los resultados. Por otra parte, mientras más difieran las características de sus pacientes con las de los pacientes estudiados, más inseguro se sentirá de extrapolar los resultados a ellos.

El segundo aspecto se relaciona con la relevancia de los resultados reportados. Cuando decidimos tratar a un paciente es porque buscamos cambios importantes en una condición. Si, por ejemplo, en pacientes con artritis reumatoide nos conformáramos con que un medicamento sólo muestre mejoría en la VHS, probablemente nuestros pacientes nos reclamarían que lo importante para ellos es cuán inflamadas tienen las articulaciones o cuánto dolor tienen. En ese caso los autores estarían sustituyendo una medición de laboratorio por aquellas cosas que verdaderamente son importantes para los pacientes. Por tanto, no sólo debemos fijarnos en que los resultados sean válidos con estimaciones de efecto confiables, sino que también se hayan considerado aquellas respuestas que son de mayor relevancia para la situación de nuestros pacientes.

Por último, debemos mirar también a la relación beneficios/riesgos, así como también a la relación costo/beneficio. Una manera de enfrentar

este aspecto es considerando el concepto de número necesario a tratar (NNT).

El NNT se refiere al número de pacientes que deben ser tratados para prevenir un evento desfavorable (o producir un evento favorable – mejoría o cura).

Supongamos que el tratamiento que se investiga produce una RAR de 5% (0,05); eso significa que por cada 20 individuos que tratemos seremos capaces de prevenir 1 evento. Si la RAR fuera de 20% (0,2) bastaría tratar 5 individuos para prevenir un evento.

Para obtener una estimación de la relación beneficios/riesgos se debe relacionar el NNT con la frecuencia de los eventos adversos. Si, por ejemplo, el estudio que analizamos demuestra una RAR de 5%, esto es, NNT de 20, y sabemos que en un 25% de los individuos tratados se produjo diarrea y en un 10% cefalea importante, entonces sabemos que:

- 1) por cada 20 tratados se evitará un evento negativo,
- 2) 5 pacientes tendrán diarrea y
- 3) 2 pacientes sufrirán de cefalea.

Respecto de la relación costo/beneficio, de manera similar a lo anterior, se relaciona el NNT con los costos del tratamiento. De esa manera, sabiendo lo que cuesta el tratamiento de un paciente, y multiplicándolo por el NNT, se puede establecer el costo de prevenir un evento desfavorable.

Todo lo anterior nos permite poner en mejor perspectiva global los resultados, sobre todo para quienes tienen que tomar decisiones administrativas para grupos de pacientes, donde además de los beneficios pueden tener una idea general de los potenciales riesgos y costos involucrados.

La intención de este artículo ha sido presentarles algunas herramientas que les ayuden a enfrentar de manera más sencilla la difícil tarea de mantenernos al día en aspectos relativos a la prevención y tratamiento de las enfermedades reumatológicas con las que lidiamos a diario.

REFERENCIAS BIBLIOGRAFICAS

1. Attard AR, Corlett MJ, Kidner NJ, Leslie AP, Fraser IA. Safety of early pain relief for acute abdominal pain. *BMJ* 1992 (Sep 5); 305(6853):554-6.
2. Pace S y Burke TF. Intravenous morphine for early pain relief in patients with acute abdominal pain. *Acad Emerg Med* 1996; 3:1086-92.
3. Jackson H. Effect of eye pads in healing of simple corneal abrasions. *BMJ* 1960; 2:713.
4. Kaiser PK. A comparison of pressure patching versus no patching for corneal abrasions due to trauma of foreign body removal. *Ophthalmology* 1995; 102:1936-42.
5. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met?. *Ann Int Med* 1985; 103:596-599.
6. www.nettingtheevidence.org.uk
7. Guyatt G, Sackett D, Cook D et al. Users' Guides to the Medical Literature: I How to use an article about therapy or prevention. A. Are the results of the study valid? *JAMA* 1993; 270 (21):2598-2601.
8. Bombardier C, Laine L, Reicin A et al. Comparison of Upper Gastrointestinal Toxicity of Rofecoxib and Naproxen in Patients with Rheumatoid Arthritis. *N Engl J Med* 2000; 343:1520-8.
9. Guyatt G, Sackett D, Cook D et al. Users' Guides to the Medical Literature. II How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients. *JAMA* 1994; 271(1):59-63.