

Interpretación de los resultados de un ensayo clínico o metaanálisis en angiología y cirugía vascular: *odds ratios*, riesgo relativo, intervalos de confianza y credibilidad

J. Escrig-Sos, J. Molina-Martínez

INTERPRETACIÓN DE LOS RESULTADOS DE UN ENSAYO CLÍNICO O METAANÁLISIS EN ANGIOLOGÍA Y CIRUGÍA VASCULAR: ODDS RATIOS, RIESGO RELATIVO, INTERVALOS DE CONFIANZA Y CREDIBILIDAD

Resumen. Objetivo y desarrollo. Artículo de divulgación en el que se explican algunas reseñas básicas para la interpretación de los resultados que aparecen en un metaanálisis o un ensayo clínico, de un modo no formal desde el punto de vista estadístico, con el fin de facilitar su comprensión para el lector poco versado en esta materia. Se acompañan de ejemplos extraídos de varios artículos en angiología y cirugía vascular, de reciente publicación, que confrontan los resultados de la angioplastia y la endarterectomía para las estenosis críticas de la arteria carótida. [ANGIOLOGÍA 2007; 59: 3-11]

Palabras clave. Ensayo clínico. Interpretación de resultados. Intervalos de confianza. Medidas de riesgo. Metaanálisis.

Introducción

Un metaanálisis o un ensayo clínico ocupan los escalafones más altos dentro de la escala de valores de la medicina basada en la evidencia. Sin embargo, esto no quiere decir que sólo por ello ofrezcan una seguridad absoluta en cuanto a la resolución de una determinada incertidumbre, ni que su lectura e interpretación sean fáciles para el clínico. En efecto, esta clase de estudios puede tener lagunas importantes en su diseño y ejecución; además, el lenguaje estadístico puede sobrepasar la capacidad de una buena comprensión para quien no se ha preocupado de adquirir, al menos, un nivel medio en esta materia, lo cual no

es precisamente habitual en la comunidad de los profesionales de la medicina.

El objetivo de este artículo es aclarar este segundo punto, de una forma que sea comprensible para los no expertos en medidas estadísticas del riesgo, utilizando un lenguaje que quizás podría considerarse como demasiado simplista por un auténtico experto, pero que, a la postre, puede resultar muy útil para quien no lo es. No entraremos en cuestiones específicas de diseño, y daremos por sentado, de entrada, que tal diseño y la ejecución, por lo menos, son suficientemente correctos. En general, los ensayos clínicos que se publican en revistas de alto prestigio suelen ser así. No obstante, el lector que tenga ante sí un estudio de estas características puede mantener a su lado un resumen de las llamadas normas consolidadas para la publicación de ensayos clínicos (CONSORT, *consolidated standards of reporting trials*) [1], que podrá orientarle sobre posibles problemas metodológicos, de una forma ordenada.

Aceptado tras revisión externa: 06.02.07.

Servicio de Cirugía General y del Aparato Digestivo. Hospital General de Castellón. Castellón de la Plana, España.

Correspondencia: Dr. Javier Escrig Sos. Servicio de Cirugía General y del Aparato Digestivo. Hospital General de Castellón. Avda. Benicàssim, s/n, 5.º B, Secretaría. E-12004 Castellón de la Plana. E-mail: escrig_vicsos@gva.es

© 2007, ANGIOLOGÍA

Aquí nos centraremos sobre el análisis de los resultados, más concretamente sobre el resultado principal que deriva del objetivo fundamental de un estudio y, por ello, sobre su diseño. Cualquier metaanálisis o ensayo clínico tiene otros resultados llamados secundarios. Hay mucha controversia sobre su utilidad y especialmente sobre su fiabilidad [2]. Si uno no es experto en metodología, es conveniente que no dé a estos resultados una gran credibilidad. En cualquier caso, es mejor que se consideren como materia de ulteriores investigaciones dirigidas especialmente a ese punto concreto o, a lo sumo, como una tendencia no demostrada todavía.

Es bueno que cualquier explicación sobre medidas estadísticas dirigidas a los no estadísticos se complemente con ejemplos prácticos. Utilizaremos tres estudios que versan sobre la comparación de la angioplastia endoluminal de la arteria carótida frente a la endarterectomía clásica. Se trata de un metaanálisis publicado en el año 2005 [3], y dos ulteriores ensayos clínicos aparecidos en el año 2006 [4,5].

Odds ratios y riesgos relativos

Así como el valor *p* indica solamente asociación entre dos variables, tanto una *odds ratio* (OR) como un riesgo relativo (RR) indican intensidad de asociación y, de forma general y desde el punto de vista epidemiológico, hablan de la intensidad de relación entre un factor de riesgo y un resultado supuestamente derivado de la actuación de dicho factor de riesgo. En nuestro caso, entre unas determinadas intervenciones sobre una estenosis carotídea y la aparición de muerte o accidente cerebral; es decir, entre factor de riesgo y enfermedad (Fig. 1).

Pero hay que distinguir algún matiz entre ellas, ya que en el plano de lo concreto no expresan lo dicho

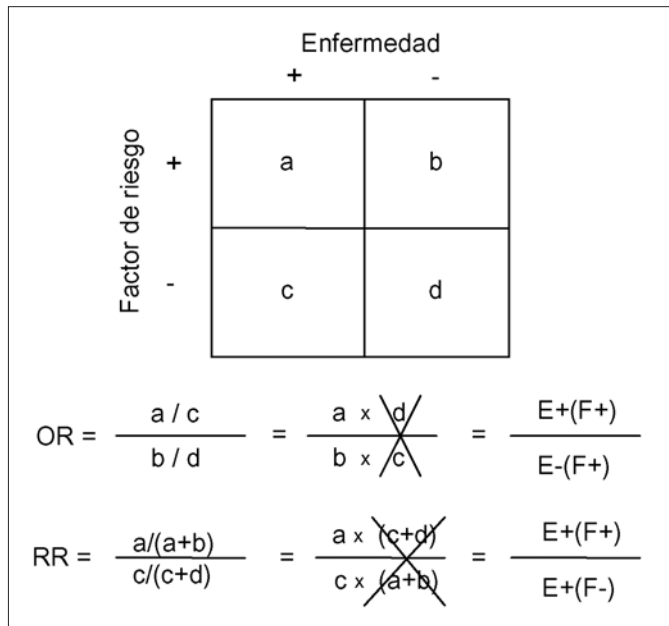


Figura 1. Fundamento matemático de una *odds ratio* y un riesgo relativo.

anteriormente de la misma forma. Si nos fijamos en la figura 1, lo primero que llama la atención es la dirección que sigue la obtención de cada medida. Una OR se calcula de forma 'vertical' dentro de la tabla, mientras que un RR sigue un sentido 'horizontal'. En dicha figura aparece tachado lo que no es más que el factor de ponderación matemático dentro de la fórmula de cada cual, y queda libre lo que representa el mensaje inmanente de cada medida. Así, una OR confronta en esencia el recuadro 'a' frente al recuadro 'b', es decir, enfermos frente a sanos sobre los que ha actuado el factor de riesgo E+(F+) frente a E-(F+), que es lo que representa, respectivamente, 'a' y 'b'. Se trata pues de una relación de opuestos en cuanto al resultado. Por el contrario, un RR es en esencia un cociente entre 'a' y 'c', es decir, entre enfermos en los que ha actuado el factor de riesgo, y de nuevo enfermos pero sobre los que no ha actuado dicho factor, expresado, respectivamente, en la figura como E+(F+) y E+(F-).

Observemos un ejemplo imaginario en la figura 2, en la que el factor de riesgo sería la angioplastia, o

| | Enfermedad | |
|-----------------|------------|----|
| | + | - |
| Angioplastia | 35 | 65 |
| Endarterectomía | 10 | 90 |

$OR = (35 \times 90) / (65 \times 10) = 4,85$
 $RR = (35 / (35 + 65)) / (10 / (10 + 90)) = 3,50$

Figura 2. Ejemplo imaginario de cálculo de una *odds ratio* y un riesgo relativo.

grupo experimental, constituyendo el grupo control la endarterectomía clásica. Si se diera el resultado que expresa la tabla, la OR valdría 4,85, y el RR, 3,50. El riesgo de enfermedad, sea embolia cerebral, infarto de miocardio, etc., sería 4,85 veces mayor para la angioplastia que para la endarterectomía, según la OR, y 3,5 veces mayor, según el RR. Esta es la forma general de lectura y, como se puede comprobar, se utiliza la apostilla 'veces más' o 'veces mayor'. Pero, de acuerdo con los matices antes señalados, la lectura más ajustada al mensaje de cada medida sería, para la OR, que hay 4,85 veces más enfermos que sanos con la angioplastia –tras ponderar según los efectos de la endarterectomía–, y para el RR que entre los enfermos, la angioplastia está presente 3,5 veces más que la endarterectomía –tras ponderar por los casos en que se aplicó cada técnica–.

Repetimos, ambas medidas expresarían que la angioplastia tendría más riesgo de resultado patológico que la endarterectomía; pero el matiz es importante. Esto es así porque una OR es una medida más imperfecta, o sólo aproximada, de lo que es la verdadera actuación de un factor de riesgo, expresada por un RR. Fijémonos que la OR tiene en cuenta, ante to-

do, el resultado de enfermos y sanos, ya pasada la actuación del factor de riesgo; por ello, es una medida que se acopla mejor para estudios retrospectivos. Sin embargo, el RR antepone primero la presencia o ausencia del factor de riesgo, para después ver qué resultado produce; así, precisa de un control estricto de la actuación de dicho factor, que sólo puede lograrse en estudios prospectivos. La importancia epidemiológica de un RR es así mayor que la de una OR, siendo ésta sólo una buena aproximación al RR, especialmente cuando no es posible hacer un diseño prospectivo, como es el caso de un estudio de casos y controles; incluso de un metaanálisis, en el que tampoco es posible reunir prospectivamente la gran cantidad de pacientes estudiados en varios ensayos clínicos. Cuando estas medidas provengan de técnicas estadísticas multivariantes, la OR es propia de la regresión logística, mientras que el RR lo es de la regresión de Cox. En el análisis de supervivencia, donde hay casos censurados o perdidos, el RR tiene un equivalente en el llamado *hazard ratio*.

El valor nulo de estas medidas es el 1. Este valor supone igualdad entre los efectos de la presencia y ausencia del factor de riesgo, o entre dos variables. Equivale pues al 0 para una diferencia entre porcentajes o medias. Así, si son mayores que 1 expresan un efecto real del factor de riesgo de cara a la enfermedad, mientras que si son menores que 1 expresarían un efecto protector de dicho factor. Siempre se ha de tener en cuenta que se están confrontando dos variables en forma de cociente; de modo que, para saber la magnitud cruda de un efecto, hay que conocer cuál es la incidencia de la enfermedad en el grupo control. La repercusión práctica de una *ratio* tres veces mayor sobre una tasa basal de infecciones de 1 por 10.000 no es igual que sobre una tasa del 15%.

Intervalos de confianza

Cualquier resultado estadístico puede expresarse

acompañado de su intervalo de confianza (IC). En el caso de una diferencia de porcentajes o medias entre dos variables, o en el caso de medidas de riesgo como la OR y el RR, su importancia es crucial para la interpretación del resultado, y puede perfectamente sustituir al valor p ; al menos, son necesarios para complementar dicho valor [6]. En términos prácticos, un IC acota aquellos valores de un resultado que son posibles en la población, alrededor de su estimación puntual en la muestra estudiada, y de acuerdo con el estudio de que se trate, así como con cierto margen de seguridad frente al error aleatorio, que suele ser del 90, 95 –el más habitual– o 99%. Para interpretar el resultado de un estudio en cuanto a su importancia práctica son insustituibles, dado que el valor p jamás indica si un resultado es más o menos importante en estos términos prácticos. En el caso de escrutar una equivalencia clínica entre dos variables, entonces no tienen otra alternativa, ya que aquí el valor p de las técnicas estadísticas convencionales simplemente no sirve para tal finalidad.

Cuando un IC contiene el valor nulo, bien el 1 para una OR o un RR, bien el 0 para una diferencia de medias o proporciones, equivale a un valor p estadísticamente no significativo. Por el contrario, si no contiene ese valor nulo, la p mostrará significación estadística. Pero, sea como sea el valor p , un IC puede contener valores intrascendentes desde el punto de vista clínico o, al contrario, tener valores de mucha trascendencia real. Si las complicaciones cerebrales o la muerte tras una angioplastia carotídea se han estimado [4] en un 3,4%, y si 1,7 y 6,7 son los límites del IC, es posible que la estimación puntual o el límite inferior de dicho intervalo no tengan trascendencia práctica; pero, si ese 6,7% o sus cifras cercanas en el intervalo son realmente importantes, el resultado del estudio también lo será, aparte de la presencia de significación estadística o de su ausencia. Así pues, de un resultado hay que mirar su estimación puntual ante todo; después, hay que disponer de su IC y ver si entre sus límites todos los valores

son trascendentes clínicamente; por último, valorar su significación estadística con el valor p , u observar si el IC contiene o no el valor nulo. En ocasiones, puede surgir un problema a tener en cuenta: cuando la muestra es excesivamente pequeña, el IC es muy ancho; dicho de otra forma, es muy impreciso. Hay que tener la precaución de no considerar como útil para sacar conclusiones un IC con límites muy distanciados entre sí. Suele ser el reflejo de un cálculo inadecuado del tamaño de la muestra. En definitiva, se trata de no dar todo el peso del mundo a la técnica estadística; más bien hay que apoyarse en ella sobre la base de nuestro buen juicio clínico.

El problema de la equivalencia y de la no inferioridad

La mayoría de ensayos clínicos tienen como objetivo valorar la superioridad de una actuación respecto a otra. En otras ocasiones –muchas menos–, los investigadores dan por sentado de antemano que un nuevo tratamiento no puede ser superior a otro clásico en cuanto a sus efectos fundamentales; pero, puede tener otras ventajas, generalmente de índole práctica en cuanto su aplicación o para ciertos efectos secundarios que, ante la equivalencia en lo fundamental, pueden aconsejar su uso de forma preferente. En este caso se plantean estudios de equivalencia absoluta o, como mínimo, de no inferioridad. Para ello, es imprescindible la lectura de forma apropiada de un IC [7]. Lo comprenderemos siguiendo la figura 3.

La bioequivalencia absoluta y la no inferioridad no son conceptos estadísticos. Se basan en el establecimiento de unos límites en cuanto al efecto estudiado, que delimitan lo que es intrascendente, de acuerdo con los conocimientos clínicos de lo que es una diferencia tolerable en ese sentido, bien en forma de diferencia de porcentajes o medias, o bien en forma de intensidad de riesgo (OR o RR). Estos límites hay que establecerlos de forma clara y de antemano, y lógica-

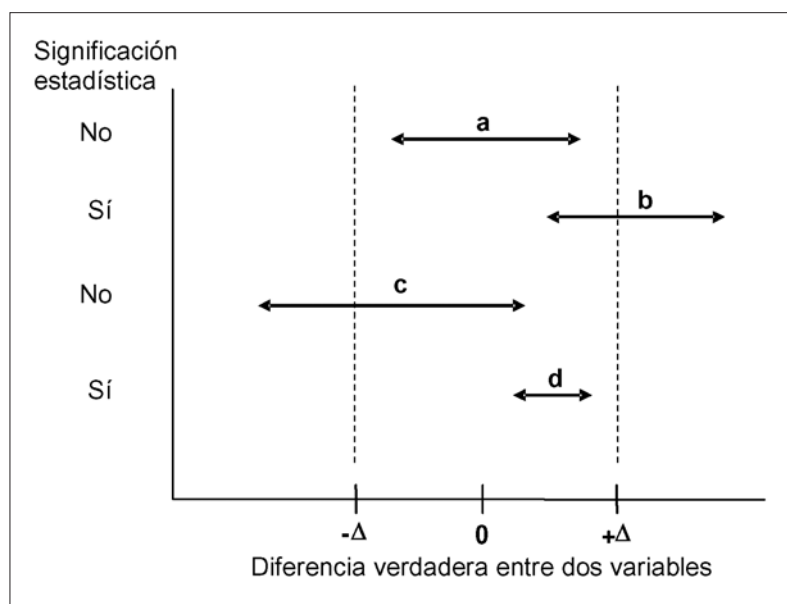


Figura 3. Equivalencia y significación estadística (modificado de [7]).

mente pueden ser motivo de polémica; pero hay que establecerlos con precisión. Es lo que en la figura 3 definen los valores representados por la letra delta (Δ). Siempre pivotarán, como es natural, alrededor del valor nulo. En el caso de equivalencia absoluta hay que definir dos límites alrededor del valor nulo y en la no inferioridad hay que establecer sólo el límite superior, aquel que representa la mayor diferencia tolerable en ese sentido, de la variable experimental frente a la que representa al grupo control.

Miremos detenidamente en la misma figura 3 las flechas que indican diversos IC para una diferencia entre variables. El intervalo 'a' sería estadísticamente no significativo, ya que contiene el valor nulo; mostraría equivalencia absoluta, ya que cae entre ambas deltas, así como no inferioridad, al no sobrepasar la delta positiva. El 'b' sería estadísticamente significativo, no equivalente, y tampoco no inferior. El 'c' no alcanzaría la significación estadística y tampoco mostraría equivalencia completa, ya que contiene el límite delta inferior; pero alcanzaría al menos la no inferioridad, al no llegar a la delta superior. El 'd' sería estadísticamente significativo al no con-

tener el valor nulo, equivalente de forma absoluta al caer entre las dos deltas, así como no inferior. Queda claro que el concepto de significación estadística en sentido clásico no vale para escrutar equivalencia absoluta ni no inferioridad, y que hay que apoyarse en los IC para ello. Si en este tipo de ensayos observamos algún valor p acompañando al IC, esta p no deriva de las pruebas clásicas de hipótesis más conocidas, sino de alguna prueba estadística especial desarrollada exclusivamente para estudios de equivalencia. No se debe olvidar que esta clase de pruebas no figuran en ciertos paquetes estadísticos

de uso más corriente, como es el caso del famoso SPSS[®] (*Statistical Product and Service Solutions*).

La credibilidad

La credibilidad del resultado de un ensayo clínico o de un metaanálisis supone también otro juicio de valor, de valor clínico y de actitud crítica, por parte del lector de un estudio. Alguien dijo una vez que la realidad supera siempre a la ficción, y este aserto también es aplicable al tema que nos ocupa, siempre que aceptemos que cualquier estudio, por perfecto que sea, no puede abarcar toda la realidad de un problema, ya que parte del análisis de una muestra y no de toda la población. La estimación puntual, el resultado de un estudio, puede ser aproximado a esa realidad, pero no igual. Si mediante ciertas técnicas estadísticas de tipo bayesiano nos aproximamos a lo que, como mínimo, debería ser el resultado real de toda la población que fuera compatible con el resultado del estudio, sólo nos quedaría aplicar nuestro buen juicio clínico como concedores del problema para deter-

minar si ese mínimo real es posible o no que pueda ser cierto, para dar un toque objetivo a la cuestión de la credibilidad de un resultado.

No se trata ahora de desarrollar matemáticamente este asunto –otros ya lo han hecho [8,9]–, sino de buscar una manera rápida y fácil de resolverlo. Ante todo hay que poner la premisa de que para hacer este análisis es necesario que un resultado muestre significación estadística. De nada sirve valorar credibilidad en un resultado que no nos saca, ya desde el principio, de la incertidumbre sobre si se debe o no al azar. Después, si se cumple lo anterior, de nuevo hay que recurrir a los IC. Para los resultados expresados como OR o RR existe una página web [10] que nos calculará rápidamente y sin más problemas ese límite mínimo de la realidad compatible con el estudio. Si el resultado nos parece desmesurado, el estudio no es creíble, aunque en apariencia su diseño o ejecución sean correctos. Generalmente esconderá algún sesgo, del tipo que fuere, más o menos perceptible [11]. Hay que leer, en este caso, con mucho detenimiento, el apartado de material y método.

El ensayo de no inferioridad del grupo EVA-3S (*Endarterectomy versus Stenting in Patients with Symptomatic Severe Carotid Stenosis*) [4] nos ofrece un RR general para accidente vascular cerebral o muerte de la angioplastia frente a endarterectomía de 2,5 (IC al 95% de seguridad de 1,2 a 5,1). Con el calculador antes mencionado y de acuerdo con los límites del IC, obtenemos que, en realidad, este RR debería ser como mínimo de 2,6 para que el resultado fuera creíble; es decir, muy poca diferencia con la estimación puntual; pero, a partir de un 260%, superior a los efectos de la endarterectomía. En otras ocasiones, la diferencia entre este límite crítico y la estimación puntual puede ser mayor. Lo que habría que valorar en este ejemplo es si 2,6 o más veces, en cuanto a la aparición de estas complicaciones de la angioplastia con respecto a lo que ofrece la endarterectomía, es o no compatible con lo clínicamente plausible. En el metaanálisis de Coward et al [3], los daños sobre pa-

res craneales de la angioplastia con respecto a la endarterectomía presentan una OR de 0,13. La angioplastia, al ser la OR menor que 1, más que un factor de riesgo es un factor de protección. Su IC al 95% de seguridad va de 0,06 a 0,25. Hay significación estadística, y el límite crítico de lo creíble se situaría como mínimo en una OR de 0,77. Esto supone que la angioplastia ponderada con respecto a lo que hace la endarterectomía produciría, como mínimo, daño en los pares craneales el 77% (unas tres cuartas partes) de las veces en que no produce tal daño. Si creemos que esa diferencia en el plano real es ya exagerada y que no puede ser mayor, entonces no es creíble una OR de 0,13, que significa que la angioplastia se queda sólo en el 13% de daños en los pares craneales con respecto a cuando no los da. El juicio clínico acerca de un problema es el que se decide en última instancia con la ayuda de las herramientas estadísticas.

Ejemplos comentados

Coward et al [3] confeccionaron un metaanálisis comparando ambas técnicas, y tomaron como grupo experimental a la angioplastia endoluminal y como referencia o control a la clásica endarterectomía. Los resultados vienen en OR al tratarse de un metaanálisis. Por la misma razón, se plantean varios resultados principales. Nos fijaremos en uno de ellos. La estimación puntual del riesgo de muerte o accidente cerebral relacionados con el procedimiento es de 1,33. La angioplastia ponderada con respecto a la endarterectomía produce un 33% más veces efectos de ese tipo que cuando no los produce. Esto es la estimación puntual, pero su IC va de 0,86 a 2,04. El resultado no es estadísticamente significativo, ya que este IC contiene el valor 1. No obstante, cabe preguntarse si el resultado es relevante clínicamente. En efecto, dentro de ese IC hay cifras que pueden ser importantes: una OR de 2,04 es un riesgo de efecto que puede ser trascendente, ya que supone en términos genera-

les unos 204% más afectados que no afectados con la angioplastia, el doble aproximadamente, 204 por cada 100. Podría haber así, en el peor escenario posible, hasta un 204% más ‘enfermos que sanos’ cuando se aplica la angioplastia. Es más, el valor nulo, el 1, está mucho más cerca del límite inferior del IC que del superior. Esto quiere decir que la probabilidad general de que la angioplastia lleve a estos efectos graves es siempre mayor que aquella que no los produzca. El resultado no es estadísticamente significativo –todo podría deberse al azar–, pero el estudio ofrece una información no despreciable porque trasluce una tendencia desfavorable para la angioplastia, aunque no se haya podido demostrar fehacientemente. La no significación estadística tampoco demuestra equivalencia. Para ello, tendríamos que establecer unos límites delta y ver cómo los afecta el IC.

Mas et al, del grupo EVA-3S [4], confeccionaron un ensayo clínico de no inferioridad, de nuevo tomando como grupo experimental a la angioplastia y como referencia a la endarterectomía, sobre una muestra de estenosis sintomáticas del 60% de la luz, y analizaron el riesgo de muerte o cualquier tipo de accidente cerebral dentro de los 30 días subsiguientes al tratamiento. El límite de no inferioridad lo establecen en un 2% más de riesgo. El estudio se detuvo cuando se habían analizado 259 pacientes con cirugía y 261 con angioplastia, porque los eventos referidos al resultado principal se produjeron en 10 casos de cirugía (3,9%) frente a 25 en la angioplastia (9,6%). Ello supone una diferencia absoluta del 5,7% en contra de la angioplastia que, traducido a RR, es de $9,6 / 3,9 = 2,5$ (estadísticamente significativo: IC al 95% de seguridad desde 1,2 a 5,1) y, a su vez, en términos del llamado ‘número necesario para dañar’ (NNH, *number needed to harm*), es de 17 (1 / 5,7); es decir, un evento adicional por cada 17 angioplastias. Estas cifras se consideraron intolerables por el comité de seguridad del estudio. Se podrían argumentar algunas críticas a esa decisión. En primer lugar, cuando se detiene prematuramente un estudio,

existe un alto riesgo de sobreestimación de las diferencias. Por otra parte si, con los datos ofrecidos en la tabla 3 del artículo, calculamos el IC con el 95% de seguridad de esa diferencia del 5,7%, obtenemos unos límites del 1-10%, cifras no aportadas por los autores; pero, pese a su significación estadística y al riesgo de sobreestimación, contienen cantidades intrascendentes, que son las que rondan el límite inferior. Estrictamente, persistía la posibilidad de que con más pacientes analizados se alcanzara el objetivo de no inferioridad. Al ser un estudio detenido prematuramente, ninguno de sus resultados ofrece la suficiente solvencia; así, no cabe entrar en más detalles que no sean también especulativos.

El grupo SPACE (*Stent-Supported Percutaneous Angioplasty of the Carotid Artery versus Endarterectomy*) [5] diseñó otro ensayo de no inferioridad para pacientes con estenosis sintomática de carótida, tomando como resultado principal el accidente isquémico ipsilateral o la muerte dentro de los 30 días subsiguientes al procedimiento. Establecieron como límite delta positivo de no inferioridad una diferencia absoluta de eventos de 2,5% en contra de la angioplastia (o a favor de la endarterectomía). La tasa de tales eventos para la angioplastia fue de 6,84 frente al 6,34% de la endarterectomía, cifras ya de por sí bastante elevadas. Pero el objetivo a estudiar, es decir, la diferencia absoluta, fue pequeña, del 0,51%, con un IC al 90% de seguridad entre un -1,89 y 2,91%. No es pues una diferencia estadísticamente significativa al contener el 0, pero el límite superior del IC sobrepasa, aunque sea por muy poco, el límite de no inferioridad preestablecido; por ello, la equivalencia en ese sentido es incierta, no se ha demostrado. De nuevo, la tendencia de la angioplastia es que, a corto plazo, ofrece peores resultados, y desde luego no puede considerarse como no inferior a la endarterectomía en ese aspecto. Lo único que podemos argüir aquí es la pertinencia de ese límite delta superior, ya que, de haber considerado sólo un poco más, por ejemplo un 3% de diferencia absoluta, se hubiera

demostrado al menos no inferioridad desde el punto de vista práctico, o de relevancia clínica, de acuerdo con el diseño utilizado por los autores. También hubiese podido resultar que con un cálculo más ajustado del tamaño de la muestra, o habiendo reclutado algunos pacientes más, el IC fuera más estrecho y se hubiera podido demostrado no inferioridad, si es que los resultados hubieran seguido la misma tendencia. Se trasluce la duda de si realmente los autores o el comité de control del ensayo no fueron demasiado exigentes para con la angioplastia.

La impresión general, si se consideran en conjunto estos tres estudios, es que la angioplastia tiene mayor riesgo inmediato de complicaciones graves que la endarterectomía. Por todo ello, se aconseja que la angioplastia carotídea sólo se use por ahora dentro de ensayos clínicos debidamente controlados, o en aquellos pacientes concretos en los que es de muy alto riesgo una endarterectomía, pero no como técnica habitual. Pero, también dejan la duda de si esas diferencias son realmente insalvables o carecen de importancia práctica verdadera. Es posible que la clave la pudiera

aportar en el futuro unos métodos de protección cerebral más desarrollados o perfeccionados.

Conclusiones

Hay que conocer el mensaje fundamental de las medidas estadísticas que aparecen en los estudios de más nivel en cuanto a evidencia, al menos de modo intuitivo, ya que no es necesario tampoco saber su entramado matemático completo para poder interpretarlos en la dirección correcta. A partir de aquí, los IC son esenciales para ello, ya que nos aportan la misma información que el valor p y, además, permiten valorar la importancia clínica de un resultado y los aspectos de equivalencia y credibilidad [12]. Por último, nuestro juicio crítico basado en nuestros conocimientos sobre el problema ha de estar siempre presente cuando leamos un trabajo de investigación. La estadística es una buena herramienta de ayuda para el análisis, pero no lo es todo; desde luego, no es la razón de ser de la investigación clínica.

Bibliografía

1. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996; 276: 637-9.
2. Cook DJ, GebSKI VJ, Keech AC. Subgroup analysis in clinical trials. *Med J Aust* 2004; 180: 289-91.
3. Coward LJ, Featherstone RL, Brown MM. Safety and efficacy of endovascular treatment of carotid artery stenosis compared with carotid endarterectomy: a Cochrane systematic review of randomized evidence. *Stroke* 2005; 36: 905-11.
4. Mas JL, Chatellier G, Beyssen B, Branchereau A, Moulin T, Becquemin JP, et al, EVA-3S Investigators. Endarterectomy versus stenting in patients with symptomatic severe carotid stenosis. *N Engl J Med* 2006; 355: 1660-71.
5. The SPACE Collaborative Group. 30 day results from the SPACE trial of stent-protected angioplasty versus carotid endarterectomy in symptomatic patients: a randomized non-inferiority trial. *Lancet* 2006; 368: 1239-47.
6. International Committee of Medical Journals Editors. Uniform requirements for manuscripts submitted to biomedical journals. *BMJ* 1988; 296: 401-5.
7. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996; 313: 36-9.
8. Matthews RAJ. Methods for assessing the credibility of clinical trial outcome. *Drug Inf J* 2001; 35: 1469-78.
9. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. Chichester, UK: Wiley; 2004.
10. Pezzullo JP. Bayesian credibility analysis: an online calculator. URL: <http://members.aol.com/johnp71/bayecred.html>. [06.11.2006].
11. Escrig J. Sobre cómo analizar la credibilidad de un ensayo clínico o metaanálisis cuyo resultado principal se ofrezca en *odds ratio*, riesgo relativo, o *hazard ratio*. *Cir Esp* 2005; 78: 351-6.
12. Escrig J. Intervalos de confianza: por qué usarlos. *Cir Esp* 2007 [in press].

INTERPRETING THE RESULTS FROM A CLINICAL TRIAL OR META-ANALYSIS IN ANGIOLOGY AND VASCULAR SURGERY: ODDS RATIOS, RELATIVE RISK, CONFIDENCE INTERVALS AND CREDIBILITY

Summary. Aim and development. *To present a non-specialised paper in which we outline some of the basic features regarding the interpretation of the results that appear in a meta-analysis or clinical trial; our aim is to do this in an informal manner as far as statistics is concerned, so as to make it easier to understand for readers who are not so familiar with the subject matter. The article includes examples from a number of recently published papers on angiology and vascular surgery which compare the outcomes of angioplasties and endarterectomies performed to correct critical stenoses of the carotid artery. [ANGIOLOGÍA 2007; 59: 3-11]*

Key words. *Clinical trial. Confidence intervals. Interpretation of results. Meta-analysis. Risk measurements.*

