Artículo especial

Localizador web
Artículo 113.013

Sobre cómo analizar la credibilidad de un ensayo clínico o metaanálisis cuyo resultado principal se ofrezca en *odds ratio*, riesgo relativo o *hazard ratio*

Javier Escrig-Sos
Cirugía General y Digestiva. Hospital General de Castellón. Castellón. España.

Resumen

Más allá del tipo de diseño y del análisis estadístico aplicado, la credibilidad de un trabajo de investigación radica en la compatibilidad de su resultado con la intensidad que aceptemos que pueda tener el fenómeno estudiado desde el punto de vista biológico. Esto requiere, en última instancia, un juicio de valor. En el presente artículo se describe un procedimiento para tratar de aproximarnos objetivamente a los límites de la intensidad que, de acuerdo con los datos que se manejen, debiera tener dicho fenómeno biológico para que, sobre esta base, nuestro juicio, derivado de los conocimientos disponibles sobre el problema, le pueda adjudicar la etiqueta de credibilidad. El procedimiento es válido cuando los resultados de la investigación se den en forma de *odds ratio*, riesgo relativo o *hazard ratio*, que a pesar de su difícil interpretación, son los estadísticos probablemente más utilizados en ensayos clínicos y metaanálisis, es decir, en los estudios con mayor nivel de evidencia en cuanto a su diseño metodológico.

Palabras clave: *Credibilidad. Ensayo clínico. Metaanálisis. Intervalo crítico.*

ON HOW TO ANALYZE THE CREDIBILITY OF A CLINICAL TRIAL OR META-ANALYSIS WHOSE MAIN RESULT IS EXPRESSED IN ODDS RATIO, RELATIVE RISK OR HAZARD RATIO

Beyond the type of design or the statistical analysis applied, the credibility of a research study lies in the compatibility of its results with the intensity that the reader could accept that the phenomenon studied might have from a biological point of view. Ultimately this requires a value judgment. The present article describes a procedure that can be used to objectively approach the limits of intensity that that a biological phenomenon could have, according to the data presented, so that, based on the reader's judgment derived from the available knowledge of the problem, the study can be deemed credible. The procedure is valid when the results of the study are expressed in odds ratio, relative risk or hazard ratio. Although these statistics are difficult to interpret, they are probably the most widely used in clinical trials and meta-analyses, that is, in studies whose methodological designs provide the highest level of evidence.

Key words: *Credibility. Clinical trial. Meta-analysis. Critical interval.*

Introducción

Estamos, sin duda, en un tiempo en el que la avalancha de información a la que tiene acceso el clínico es ingente y desproporcionada para lo que abarca ya no sólo su capacidad de análisis, sino también el tiempo disponible de lectura. El primer problema derivado de ello es poder separar el grano de la paja, y hacerlo con fundamento, tarea no siempre fácil, pues como glosó Smith[1,2], exeditor del *British Medical Journal*, el clínico no es un héroe que deba dominar a fondo las herramientas estadísticas y metodológicas de la investigación para las que no ha sido específicamente preparado, igual que pueda dominar la materia que es objeto de su trabajo diario. El segundo problema es escoger el grano que sea de calidad, que pueda verdaderamente aplicarse en su desarrollo profesional, y que sea susceptible de producir un cambio positivo en éste. Para que esto ocurra es necesario que creamos como posible aquello que se nos dice en la bibliografía. Nadie puede negar que al acabar de leer un

Correspondencia: Dr. J. Escrig Sos.
Cirugía General y Digestiva. Hospital General de Castellón.
Avda. Benicasim, s/n. 12004 Castellón. España.
Correo electrónico: escrig_vicsos@gva.es

Manuscrito recibido el 6-4-2005 y aceptado el 16-6-2005.

trabajo de investigación, en última instancia, lo que se hace es precisamente un ejercicio de credibilidad.

La medicina basada en la evidencia, con sus postulados, junto con la expansión y la accesibilidad de ciertas herramientas estadísticas, han contribuido enormemente a allanar este camino, pero al final es el propio clínico quien decide si lo que lee es realmente creíble, es decir, si es aceptable según lo que él conoce acerca del fenómeno de que se trate, y según la idea que posee sobre hasta qué punto o con qué intensidad dicho fenómeno pueda ocurrir en la naturaleza. Se trata de un paso que sobresale de lo que es un análisis estadístico pertinente, una interpretación correcta de éste, y un oportuno diseño del estudio que se analiza, etapas éstas imprescindibles pero no suficientes por sí mismas para que un resultado sea creíble. La credibilidad se basa, pues, en un juicio cualitativo, pero lleva implícito el conocimiento aproximado o la intuición, en cuanto a un resultado, de unos límites de naturaleza cuantitativa que no se pueden traspasar. La lectura crítica de un artículo se limita muchas veces a la búsqueda de fallos groseros en éste, mientras que el aspecto de su credibilidad se suele dar por resuelto tan sólo según el valor de la p o según el propio diseño del trabajo –lo que es un abordaje preliminar del problema, aunque es también insuficiente–, o ni se repara en él, puesto que la mayor parte de las veces ni se le hace mención en la discusión. Es preciso tener, pues, grandes conocimientos metodológicos y del problema investigado, o de lo contrario recurrir a comentarios que aparecen tras la publicación del estudio en cuestión para conformar una idea más concreta de su credibilidad. De hecho, las revistas médicas deberían aumentar más de lo que lo hacen ya la proporción de artículos de revisión y divulgación[2].

En cirugía, afortunadamente, en los últimos años se ha multiplicado la aparición de ensayos clínicos y metaanálisis, cuya escasez años atrás se nos echaba en cara por parte de otros colegas. Posiblemente la aparición de la cirugía laparoscópica haya sido el gran estímulo. En bastantes de ellos se desarrolla una parafernalia estadística ciertamente sofisticada; muchos presentan un diseño impecable, pero como en nuestro campo se manejan sobre todo variables cualitativas y predominan ciertas pruebas estadísticas, es frecuente expresar los resultados en forma de razón *(ratio)*, medidas que son de interpretación bastante críptica y muy poco intuitiva o inmediata[3]. En este artículo se intentará dar unas orientaciones acerca de cómo interpretar de forma rápida, correcta y con pocas matemáticas, unas medidas de carácter cualitativo muy frecuentes en bibliografía médica, sobre todo en estudios de alto nivel, como ensayos clínicos y metaanálisis, como son las expresadas en forma de razones o ratios: *odds ratio* (OR), *riesgo relativo* (RR), o lo que es igual cuando se trata de estudios de supervivencia, *hazard ratio* (HR), para después mostrar un sistema relativamente sencillo de aproximarse a los límites cuantitativos que debe presentar un fenómeno, para que, sobre la base de lo expresado en el estudio que estamos leyendo, podamos juzgar si éste es creíble en cuanto a sus resultados, dando por sentado que el diseño de tal estudio sea correcto y las herramientas estadísticas usadas sean las apropiadas y muestren un valor de p suficientemente pequeño que permita aceptar que el efecto que se comunica existe realmente y en la dirección en que se informa. Todo ello se apoyará sobre ejemplos de estudios, alguno muy conocido, de reciente aparición en el campo de la cirugía general.

Cómo se lee una OR y medidas similares

Aunque en su fórmula matemática difieran, pues ésta debe corresponderse con las distintas situaciones en que se utilizan, la lectura de una OR, un RR y un HR es idéntica. Basta aplicar las palabras “veces más” o “por cada 1”. Si en la comparación en cuanto a curaciones de un novedoso tratamiento A frente a un antiguo tratamiento B se produce una OR de 1,3, significa que A cura 1,3 veces más que B, o en A se producen 1,3 curaciones por cada 1 curación que en B. Por tanto, A cura más que B. Siempre hay un grupo al que se le endosa la *ratio*, que es el grupo que se considera experimental (en este ejemplo es A), y otro grupo que sirve de referencia o de grupo control, al que se le endosa el 1, cifra que ni siquiera se cita en el resultado, que en este ejemplo es el grupo B. Si se tratara de un estudio de supervivencia, como en una regresión de Cox, el HR (algunos aquí también lo llaman RR) indicaría que el riesgo de eventos terminales que se podrían esperar en A durante el tiempo de seguimiento es 1,3 veces más alta que en B; por tanto, la supervivencia sería 1,3 veces peor en A que en B.

Si, a pesar de que no se exprese, el riesgo del grupo control se enrasa a 1, ello quiere decir que este 1 es el cero de la escala. Así, una OR, un RR o un HR de 1 significa que en A se produce el mismo efecto que en B, pues 1 evento en A por cada 1 en B supone un empate. Como ya habrán observado en la bibliografía, este tipo de estadístico se suele reportar, cuando se refiere al resultado principal del estudio, acompañado del llamado intervalo de confianza (IC), que es expresión de lo que la variabilidad debida al azar puede hacerlo oscilar en la población, o en la realidad. De esta forma, lo que en un estudio se calcula y se ofrece como un resultado puntual, no tiene por qué corresponderse fielmente con el verdadero valor en la población, valor que siempre será desconocido; pero el IC, a fin de cuentas y a pesar de que matemáticamente no sea exactamente así, expresa los límites que con una cierta seguridad contendrán ese verdadero valor. Generalmente los IC se calculan con una seguridad del 95%, lo que quiere constatar que, a la postre, sólo deja una probabilidad de alrededor de 0,05 de que el verdadero valor del OR no se halle en este intervalo. Esto es fácil de traducir aproximadamente a lo que es el valor p. Si el IC del 95% de una *ratio* no contiene el 1, el grado de evidencia de que los datos del estudio sean compatibles con la hipótesis nula –que es expresión de igualdad entre lo que se compara– ha de ser menor de 0,05, y viceversa. Los IC, de este modo, nos proporcionan la misma información cualitativa que el valor de p, pero además nos dan también una idea inicial de la intensidad posible del fenómeno estudiado a través de sus propios límites, siempre según los datos de la muestra empleada, idea que tiene que ver con lo que se podría llamar importancia práctica de un resultado, lo cual sin

duda es otro paso inicial más para valorar la credibilidad de éste.

La OR o similares deben ser, pues, estadísticamente distintos de 1 para aceptar con cierta garantía que existe un efecto, es decir, para marcarlos con la famosa etiqueta de estadísticamente significativos. Antes hemos puesto el ejemplo del valor de 1,3, pero imaginemos que en lugar de 1,3 el resultado hubiese sido menor que 1, por ejemplo 0,2. Si B sigue considerándose como grupo de referencia y A como el experimental, a B hay que seguir endosándole el 1. Por tanto, decir que A produce 0,2 curaciones más que B, o que produce 0,2 contra 1 curación en B, es como decir que A cura menos que B. Un valor menor que 1 indica, pues, un efecto menor para el grupo experimental, y viceversa. Si el IC del 95% de seguridad de 0,2 se reportara, por ejemplo, como de 0,1 a 0,5, el resultado será estadísticamente distinto de la igualdad ($p < 0,05$) puesto que este IC no contiene el 1. Si los límites estuvieran entre 0,06 y 1,2 no existiría tal significación estadística, pues el 1 cae dentro del intervalo.

Aunque lo anterior no sea excesivamente difícil de comprender ni de recordar, el verdadero problema de esta clase de estadístico, tan frecuente en ensayos clínicos y metaanálisis, es que no da una idea inmediata de la magnitud absoluta de la diferencia entre los grupos comparados. Si en el estudio analizado no se refiere el porcentaje absoluto de curaciones que suceden en B, desde el valor de la OR no podremos saber las curaciones que pueden ocurrir en A, ni la mencionada diferencia cruda entre ambas; por tanto, no podremos realizar un juicio de valor sobre el alcance del fenómeno de una forma más asequible a su comprensión. Téngase en cuenta que, por ejemplo, doblar una tasa de infecciones puede resultar dramático si en el grupo control hay un 10%, pero puede ser intrascendente si se dobla un 1 por mil.

Existe una sencilla fórmula para empezar a resolver esta oscuridad inherente a este tipo de medidas, denominada diferencia relativa de riesgo[3] (DRR): cuando una OR, un RR o un HR sea mayor que 1 se aplicará esta fórmula, cuyo resultado es una proporción:

$$DRR = OR - 1 \text{ (fórmula 1).}$$

Cuando una OR sea menor que 1 se aplicará esta variante:

$$DRR = 1 - OR \text{ (fórmula 2).}$$

Si una OR de 0,2 indicaba que en A ocurrían menos curaciones que en B, al aplicar la fórmula 2, se nos proporciona un resultado para la DRR de 0,8, que nos traduce que en A se produce el 80% de las curaciones que proporciona el tratamiento B, lo que es como decir que en B se da un 20% de curaciones más que en A. Si la OR fuera de 1,3, según la fórmula 1, se nos está indicando directamente que en A se da un 30% más de las curaciones que en B. Aun a pesar de calcular estas DRR, a efectos prácticos, todavía estas cifras no pueden considerarse altas ni bajas, si no recordamos lo que se decía antes, ya que todo dependerá de lo que sea la proporción absoluta de curaciones en el grupo control (B), para valorar como grande o pequeña una diferencia que es relativa, sea del porcentaje que sea, y para poder afirmar consecuentemente que supone una diferencia clínica absoluta importante, o no. Nunca olvidemos, pues, que para valorar de forma completa el mensaje de una OR, un RR o un HR, hay que escrutar entre los datos y las tablas que nos ofrecen los autores, qué proporción absoluta de eventos se producen en el grupo control o de referencia, ayudándonos también del simple cálculo de la diferencia relativa de riesgo, que no es más que una traducción de las *ratio* a proporciones o porcentajes, medida ésta a la que estamos mentalmente más acostumbrados a interpretar.

Cálculo del intervalo crítico

La mayoría de los estudios trabajan sobre muestras más o menos grandes que proceden de una población. La auténtica realidad de un fenómeno sólo puede alcanzarse analizando la totalidad de esa población, pero es inalcanzable en toda su amplitud para cualquier tipo de análisis y diseño a partir de una muestra, por grande que sea. Para que un estudio sea creíble, su resultado ha de quedar englobado dentro de un rango de valores determinado que sean posibles dentro de lo que es la realidad biológica del fenómeno estudiado. Debe existir, pues, lo que llamaremos intervalo crítico (*critical prior interval*[4,5] en su acepción más técnica) y, además, si aceptamos que el fenómeno existe, ha de haber una diferencia entre lo que se compara, que a su vez debe presentar una determinada dirección compatible con la realidad. Esto se podrá cumplir sólo si los individuos estudiados tuviesen de entrada la misma probabilidad de ser sometidos a las situaciones que se comparan, es decir, realmente sólo valen estudios aleatorizados (lo demás son aproximaciones), y si esa dirección puede ser aceptada además de biológicamente también estadísticamente, es decir, si el IC de una *ratio* no contiene el 1 y puede ser considerado como estadísticamente significativo en un determinado nivel.

De este modo, y basándose en estos principios, es posible calcular siempre a partir del IC de las *ratio* que se ofrecen en un estudio, otro intervalo crítico que, alejándose desde el valor 1 en dirección ascendente o descendente, llegue hasta una cierta cifra límite (límite crítico) desde la cual sea posible que, en cualquier circunstancia, el resultado del estudio tenga que seguir mostrando significación estadística y en el mismo sentido que indica tal resultado. Conocido este límite, sobre él se debe aplicar el juicio clínico personal para aceptar o rechazar que el fenómeno analizado puede llegar a ser en la realidad tan intenso como lo define este límite. Si excede lo creíble de acuerdo con nuestros conocimientos sobre el problema, no podremos corroborar la credibilidad de un estudio, por más pequeño que sea el valor p, y por más impecable que parezca su diseño y su ejecución.

La falta meridiana de credibilidad definida de esta forma es un gran indicador de sesgo. Cuando aparece es necesario que algún tipo de sesgo se haya entrometido en el estudio, sin mencionar la palabra *trampa*, porque somos bien pensados. Lo que ocurre es que un sesgo, sea del tipo que sea, no siempre se podrá demostrar ni siquiera desde la lectura detenida de un trabajo científico,

aunque es posible que algún tipo de sospecha, como veremos en los ejemplos, se pueda tener.

Resulta sorprendente que un razonamiento como el expuesto no haya tenido calado hasta la fecha en la bibliografía médica, a pesar de la medicina basada en la evidencia, que se ha quedado justo en su antesala, cuando ha sido bastante utilizado en otras ramas del saber. Su idea y su desarrollo para la aplicación médica sobre los resultados expresados en *ratio* se debe a Matthews[4], un físico matemático que también se ocupa de temas metodológicos, y su introducción y su aplicación en la investigación clínica se debe a la mención que de él se hace en el libro de Spiegelhalter et al[5], conocido bioestadístico del Medical Research Council, que versa sobre el enfoque bayesiano de los ensayos clínicos. La fórmula que calcula este límite crítico, para una OR o similares menor que 1, es la siguiente (fórmula 3):

$$L_0 = \exp\left\{ - \frac{[\ln(U_D/L_D)]^2}{4\sqrt{\ln(U_D)\ln(L_D)}} \right\}$$

siendo: ln logaritmo natural; U_D el límite superior del IC del resultado expresado en *ratio* y L_D su límite inferior; exp significa exponencial. Cuando una OR o similares sean mayores que 1, se aplica la misma fórmula pero en lugar de U_D y L_D se deben introducir sus inversos: $1/U_D$ y $1/L_D$, respectivamente (fórmula 4).

Estas fórmulas son relativamente sencillas de implementar en una hoja de cálculo si uno se maneja tan sólo como usuario; no obstante, en el anexo final se ofrece un nomograma (fig. 1) que facilita su cálculo inmediato. También existe una página en internet que proporciona una explicación del procedimiento y su cálculo automático: http://members.aol.com/johnp71/bayecred.html

Ejemplo 1

Lacy et al[6] publicaron un ensayo clínico que comparaba la cirugía laparoscópica con la convencional en el cáncer de colon no metastásico. El estudio aparentemente era totalmente correcto desde el punto de vista de su diseño y su ejecución, y su resultado fue espectacular, por lo que tuvo y sigue teniendo un gran eco. Al aplicar un análisis de supervivencia, entre sus resultados principales destacaba un HR de 0,38 (IC del 95% entre 0,16 y 0,91) con respecto a muertes relacionadas con el tumor durante el seguimiento, que favorece a la cirugía laparoscópica: 0,38 muertes por cada 1 de la cirugía abierta. Si atendemos a la significación estadística y a la calidad del estudio, no se pueden poner obstáculos cara a la credibilidad. Pero enfoquémosla en el sentido de calcular el límite del intervalo crítico y valorarlo a continuación.

Se obtiene, así, un límite crítico (fórmula 3) del HR de valor 0,16. Al calcular su DRR (fórmula 2) obtenemos una diferencia relativa de, al menos, un 84% a favor de la cirugía laparoscópica, es decir, para que el resultado que da el autor sea creíble, en la realidad la cirugía laparoscópica debería presentar un resultado mejor que ese límite, o sea, menos del 84% de mortalidad relacionada con el tumor, que la que tenga la cirugía abierta durante ese mismo período de seguimiento. Para interpretar este límite, casi es indiferente, sea cual sea, la cifra general de mortalidad por cáncer de colon no metastásico a largo plazo que se obtenga con cirugía abierta, pues da la sensación de que como mínimo un 16% de reducción de la mortalidad solamente gracias a la laparoscopia es una cuantía demasiado grande para poder ser real, a la luz de los conocimientos existentes, y considerando que la cirugía laparoscópica sólo cambia en principio la vía de abordaje y no la radicalidad del procedimiento de resección tumoral. Como han señalado otros autores[7] que han comentado este trabajo, los resultados con la cirugía abierta distaron de ser óptimos, bastante peores de lo que se podría esperar de un grupo de gran excelencia, frente a unos resultados muy brillantes en el grupo laparoscópico. Quizá aquí esté la clave de un resultado muy difícil de explicar biológicamente, y sobre cuya intensidad real derivada del ensayo clínico en cuestión sólo podemos obtener una valoración objetiva a través del método que proponemos en este artículo. De entrada, siempre hay que ser escépticos con ensayos clínicos con resultados espectaculares, pues incluso el azar puede alterar un diseño correcto. Ejemplos hay en la bibliografía[8]; en estos casos puede haber sucedido una violación, aunque sea inconsciente, del principio de incertidumbre de los ensayos clínicos, cosa muy difícil de detectar, por otra parte[9], de modo que incluso la propia aleatorización puede favorecer a alguno de los grupos que se comparan.

Ejemplo 2

Aunque nos saltemos una de las condiciones previas necesaria para demostrar credibilidad absoluta, con intención meramente ilustrativa del procedimiento, comprobaremos los resultados de un estudio retrospectivo no aleatorizado que compara los resultados de la linfadenectomía clásica (D1) frente a la linfadenectomía extendida (D2) en el cáncer gástrico. Sierra et al[10] publicaron los resultados de su serie en la que encontraron una *ratio* independiente de muertes debidas al tumor durante el seguimiento de 2,3 (IC del 95% entre 1,25 y 4,3) entre el grupo D1 con respecto al grupo D2, es decir, la supervivencia era 2,3 veces peor, claramente pues, en el grupo con linfadenectomía D1.

Traducido a intervalos críticos, se obtiene un límite crítico de 1,95 que, en forma de diferencia relativa, se traduce en que en los D1 es necesario que, en la realidad, se produzca al menos un aumento relativo de muertes del 95% con respecto a los D2 sólo debido al tipo de linfadenectomía, para que el resultado del estudio sea plausible. Sin negar que es posible que la linfadenectomía D2 en algún subgrupo de pacientes pudiera mejorar la supervivencia, este resultado se nos antoja demasiado abultado como para darle presunción de credibilidad.

Un análisis de una serie de casos puede contener infinidad de sesgos, especialmente de selección. Con frecuencia en este tipo de trabajos se presenta alguna tabla con datos cruciales sobre las características de los pa-

cientes de cada grupo a los que se aplica alguna prueba estadística al uso, que en caso de no demostrar diferencias estadísticamente significativas entre estas características, parece que dé carta de homogeneidad entre esos grupos, como si una aleatorización previa se hubiera llevado a cabo. En este trabajo también aparece una tabla de esta clase. Tal cosa no es más que una trampa conceptual o un autoengaño que se hacen muchos autores, pues como alguien[11] afirmó, "la ausencia de la evidencia no es evidencia de la ausencia" cuando se usan pruebas de significación. No se puede sustituir así un proceso de aleatorización. En este estudio, además, resultó que sí existían diferencias entre los grupos que llevan a sospechar que la linfadenectomía radical se aplicó preferentemente a pacientes más jóvenes, con mejor puntuación ASA, y se acompañó de intervenciones más extensas y con mayor porcentaje de administración de radioterapia. Todo esto podría explicar las diferencias tan espectaculares obtenidas.

Ejemplo 3

En una sección de una revista[12], con el título de "Cirugía Basada en la Evidencia", aparecen unos resúmenes de revisiones sistemáticas de la *Cochrane Library*, una de las cuales se refiere a la comparación entre quimioterapia intensiva y trasplante medular frente a quimioterapia convencional en el cáncer de mama de mal pronóstico. Sólo se encuentran diferencias significativas en cuanto a intervalo libre de enfermedad únicamente al tercer año de seguimiento a favor del tratamiento intensivo, con un RR de 1,11 (IC del 95% entre 1,05 y 1,18). Aplicado el cálculo del intervalo crítico (fórmula 4) se obtiene un límite crítico de 1,04. En términos de diferencia relativa de riesgo esto supone que es necesario que en la realidad haya un mínimo de un 4% más de recidivas al tercer año en el grupo de tratamiento convencional para que este resultado sea plausible con esa realidad. Esta cifra puede ser perfectamente asumida si tenemos en cuenta que este tipo de tratamientos no parecen tener grandes ventajas en estos tumores inicialmente muy avanzados, por lo tanto puede darse como creíble, y en general orienta indirectamente hacia cierta equivalencia en esos resultados por más que las diferencias fueran estadísticamente significativas.

Conclusiones

Atender sólo a un valor de p significativo es una forma muy débil de valorar la credibilidad de un resultado en el sentido de ser concordante con lo que pueda ocurrir en la realidad biológica del fenómeno estudiado[4]. Su orientación filosófica sólo va dirigida a evaluar la compatibilidad de unos datos con una hipótesis concreta que supone igualdad entre lo que se compara, no lo plausible biológicamente[13]. Por otra parte, ni el mejor diseño metodológico ni la prueba estadística más sofisticada pueden ponernos a salvo por completo de la presencia de algún tipo de sesgo.

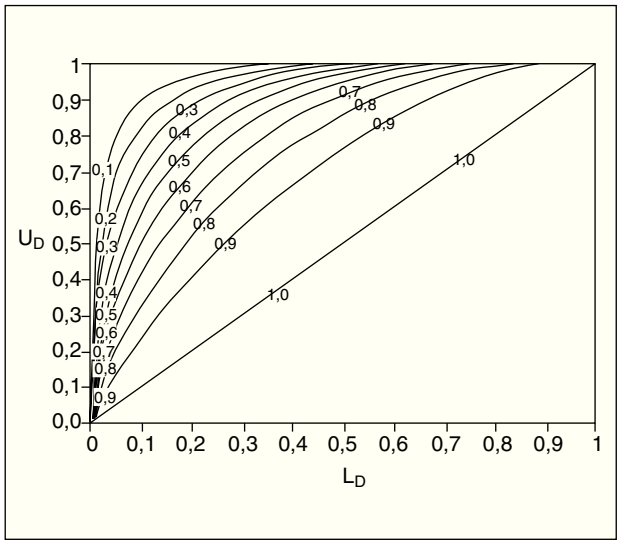

Fig. 1. Nomograma para el cálculo del intervalo crítico para odds ratio *(OR), riesgo relativo (RR) y* hazard ratio *(HR). (Reproducido de Matthews[4], con permiso del autor.)*

El método que se propone va más allá de estas limitaciones, aunque tampoco sea una panacea. Habrá situaciones en que pueda despejar las dudas sobre la credibilidad de un resultado, en un sentido u otro, pero en otras también nos puede dejar en el terreno de la incertidumbre. En cualquier caso, es una herramienta que ofrece un valor añadido en este difícil y resbaladizo terreno, y puede ser de gran ayuda para el estímulo de la lectura crítica de la biblografía, por ejemplo, entre los colegas más jóvenes cuando tengan que evaluar un trabajo de investigación en una sesión bibliográfica, o cara a la práctica clínica diaria.

Anexo

En la figura 1 se ofrece un nomograma para el cálculo del límite del intervalo crítico de una OR o de sus estadísticos similares, a partir de los IC del 95% de seguridad de éstos. Cuando una *ratio* sea menor que 1 se puede acudir directamente al nomograma, teniendo en cuenta que U_D es el límite superior de dicho IC, y L_D es el límite inferior.

En caso de que la *ratio* a valorar sea mayor que 1 hay que calcular los inversos de los límites de su IC pero teniendo en cuenta que entonces el U_D original, al transformarse en $1/U_D$, pasa a ser el L_D en el nomograma, y el L_D original, al calcular su $1/L_D$, pasará a ser el U_D. Una vez confrontados en el nomograma y obtenido así el valor límite del intervalo crítico, hemos de calcular por último su inverso para conocer la cifra que buscamos.

Como se podrá comprobar, si intentamos aplicar las fórmulas del intervalo crítico o acudir al nomograma, a partir del IC de una *ratio* que sea estadísticamente no significativa (que contenga el 1), el cálculo del intervalo crítico resulta imposible. Cuando el valor de p es muy alto, el resultado del estudio podría ser fruto del azar, y bajo esta premisa es inadecuado hablar de credibilidad, sin menoscabo de que un estudio así pueda aportar información interesante.

Bibliografía

1. Smith R. Doctors are not scientist. BMJ. 2004;328:7454.
2. Smith R. Traveling but never arriving: reflections of a retiring editor. BMJ. 2004;329:342-4.
3. Argimón JM, Jiménez J. Métodos de investigación clínica y epidemiológica. 2.ª ed. Madrid: Ediciones Harcourt; 2002.
4. Matthews RAJ. Methods for assessing the credibility of clinical trial outcome. Drug Information Journal. 2001;35:1469-78.
5. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. Chichester: Wiley; 2004.
6. Lacy AM, Garcia-Valdecasas JC, Delgado S, Castells A, Taura P, Pique JM, et al. Laparoscopy assisted colectomy versus open colectomy for treatment of nonmetastatic colon cancer: a randomised trial. Lancet. 2002;359:2224-9.
7. McLeod RS, Stern H. Evidence Based Reviews in Surgery (10). Can J Surg. 2004;47:209-11.
8. Wheatley K, Phil D, Clayton D. Be skeptical about unexpected large apparent treatment effects: the case of an MRC AML12 randomization. Controlled Clinical Trials. 2003;24:66-70.
9. Joffe S, Harrington DP, George SL, Emanuel EJ, Budzinski LA, Weeks JC. Satisfaction of the uncertainty principle in cancer clinical trials: retrospective cohort analysis. BMJ. 2004;328:1463-8.
10. Sierra A, Regueira FM, Hernández-Lizoáin JL, Pardo F, Martínez-Gonzalez MA, Álvarez-Cienfuegos J. Role of the extended lymphadenectomy in gastric cancer surgery: experience in a single Institution. Ann Surg Oncol. 2003;10:219-26.
11. Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ. 1995;311:485.
12. Anónimo. Evidence-based surgery. J Am Coll Surg. 2003;196:950.
13. Prieto L, Herranz I. ¿Qué significa estadísticamente significativo? La falacia del 5% en la investigación científica. Madrid: Díaz de Santos; 2004.