

El uso del riesgo relativo (RR), el odds ratio (OR) y la significación estadística "limítrofe"

El interesante trabajo de los Dres. C. Tajer, H. Doval, V. Mauro y colaboradores, publicado recientemente, (1) aborda desde la perspectiva del metaanálisis la comparación de dos métodos para el tratamiento del síndrome coronario agudo sin elevación del ST. La lectura atenta del trabajo permite observar que los valores referidos como RR en la Tabla 5 aparecen como OR en el resto del trabajo (Tabla del Resumen, Summary, Figuras 1, 2 y 3 y Tablas 2,3 y 4). Considerado como un error tipográfico, vale simplemente la observación; pero un análisis más profundo podría dar la oportunidad de señalar algunas diferencias entre los valores de RR, OR e intervalos de confianza respectivos, y en especial de la comúnmente llamada significación estadística "limítrofe". El uso preferencial del OR se debe a que éste proporciona un enlace entre los estudios de cohortes y los de caso-control, y a su vez con el modelo de regresión logística, principal método de análisis para los últimos. Pero las sutilezas de la estadística hacen que en algunas situaciones fronterizas, y de acuerdo con el indicador y método que se use, un resultado pueda estar de un lado u otro de la delgada línea que separa la significación estadística y en consecuencia comprometer las conclusiones de un estudio. En la Tabla adjunta se han calculado nuevamente los valores de RR y OR del trabajo, y en base a diferentes métodos (Taylor para RR y Cornfield (2) para OR, usados en el EpiInfo 5.5) los intervalos de confianza del 95% para RR y OR. Por otro lado, los mismos intervalos de confianza también se calcularon aplicando las conocidas fórmulas convencionales: (3)

Exp $[\ln RR \pm 1,96 \cdot (1-a/(a+b))/a + (1-c/(c+d))/c]$ para IC... de RR, y

Exp $[\ln RR \pm 1,96 \cdot 1/a + 1/b + 1/c + 1/d]$ para IC... de OR

siendo a, b, c y d los valores correspondientes a cada una de las celdas de una tabla de contingencia de 2 x 2 (véase tabla).

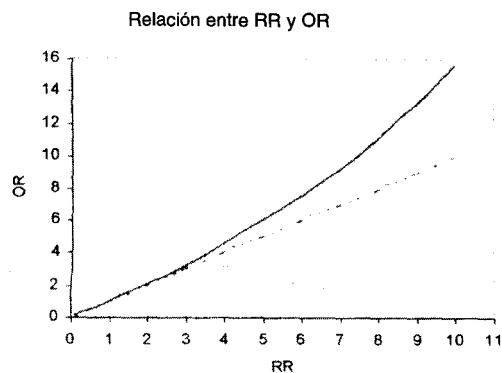
En el caso de comparaciones de varios estudios con frecuencias pequeñas, la estimación del conjunto de datos ponderados o *pooled* puede hacerse con el método de Mantel y Haenszel (4) (R... o estimación acumulada de OR) de acuerdo con:

$$R... = \frac{a \cdot d / n}{b \cdot c / n_i}$$

Mantel y Haenszel no formularon una forma de estimar la varianza de R... y en consecuencia tampoco un método para la construcción de límites de confianza. Es así que se recomiendan habitualmente otras opciones para su cálculo, a saber, los límites de confianza "basados en pruebas" de Miettinen, (5) de Breslow, (6) de Robins, (7) etc. En la última columna de la tabla se incluyeron los correspondientes valores de R... para cada serie y sus respectivos intervalos de confianza del 95% calculados con el método de Miettinen. En el ejemplo de la tabla adjunta sólo se calcularon los ítem que habían tenido significación estadística en el trabajo original, y como se observa, los valores de los intervalos de confianza del 95% varían con cada método de forma tal de hacer significativa o no la respectiva **variable en el nivel** del 5% (inclusión o no del RR = 1, OR = 1 o $\chi^2 = 1$).

Dejando de lado estas situaciones en las que los resultados pudieran encontrarse en esta zona límite

	Invasivo	Conservador	RR e IC95%	OR e IC95%	R -
MOrtalidad 30/36 días s/FRISC	66/2.316	45/2.297	1,45 Taylor (1,00-2,12) Convencional [1,01-2,13]	1,47 Cornfield (0,98-2,19) Convencional [1,00-2,15]	1,47 Miettinen (0,98-2,19)
Infarto 0 muerte a los 6-12 meses	399/3.523	453/3.523	0,88 Taylor (0,78-1,00) Convencional [0,76-0,98]	0,87 Cornfield (0,75-1,00) Convencional [0,75-1,00]	0,86 Miettinen (0,74-1,00)
Infarto a los 6-12 meses	272/3.535	336/3.531	0,81 Taylor (0,69-0,94) Convencional [0,68-0,92]	0,79 Cornfield (0,67-0,94) Convencional [0,67-0,94]	0,79 Miettinen (0,66-0,94)



para la significación estadística, podría argumentarse que en general los valores de RR y OR pueden ser tan parecidos como para ser intercambiados. En general los valores de OR suelen ser algo mayores que los de RR, pero se aproximan bastante cuando la mayor de las dos proporciones comparadas es inferior a 0,10, o sea cuando el evento es raro. En la figura (véase arriba) se ve que esto es verdad para cierto rango de valores de RR, más allá de los cuales se manifiesta un crecimiento de los valores de OR como si se tratara de una función de tipo potencial, lo que demuestra lo inapropiado de su intercambio.

En resumen, el valor de un metaanálisis se basa ciertamente en la información extra que se puede obtener con la ponderación de los trabajos en conjunto, habida cuenta del incremento en el tamaño de la muestra que se genera con este enfoque. En consecuencia, los resultados finales del metaanálisis de la Tabla 5 podrían reflejar en realidad, y de acuerdo con la técnica de análisis empleada, que "la estrategia invasiva inicial no tendría ventajas en términos de mortalidad inmediata o alejada, lo mismo en lo que respecta a la reducción de los eventos combinados (significación estadística límite), aunque sí existiría ventaja en la disminución de la incidencia de infarto en el seguimiento alejado". Sin duda, las prudentes conclusiones y las limitaciones señaladas al final del trabajo del Dr. Tajer y colaboradores resaltan la rigurosidad y honestidad científica de los autores.

Dr. Raúl A. Borracci
PROFESOR DE BIOESTADÍSTICA
UNIVERSIDAD AUSTRAL, ARGENTINA

BIBLIOGRAFIA

1. Tajer C, DOval H, Mauro V y cOl. Estrategias invasivas comparadas con estrategias conservadoras en la angina inestable y el infarto sin elevación del ST al ingreso. Metaanálisis. Rev Argent Cardiol 2001; 69: 347-361.
2. Cornfield J. A statistical property arising from retrospective studies. Proc Third Berkeley Symp Math Stat Prob 1956; 4: 135-148.
3. DawsOn-Saunders B, Trapp RG. Intervalo de confianza para el riesgo relativo y la razón desigual. En: BiOestadística médica. 2ª ed. México, El Manual ModernO. 1997; pp 202-203.

4. Mantel N, Haenszel W Statistical aspects Of the analysis Of data from retrOspective studies disease. J Nat Cancer Inst 1959; 22: 719-748.
5. Miettinen OS. Estimability and estimation in case-referent studies. Am J Epidemiol 1976; 103: 226-235.
6. Breslow NE, Liang KY. The variance of the Mantel-Haenszel estimator. Biometrics 1982; 38: 943-952.
7. Robins JM, Breslow NE, Greenland S. Estimators Of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models. Biometrics 1986; 42: 311-323.

Interpretación clínica de los conceptos estadísticos

¿Odds ratio o riesgo relativo? y la "ilusión" de la $p < 0,05$ como criterio de significación estadística

Agradecemos el interés y los comentarios del Dr. Raúl Borracci a nuestro metaanálisis de publicación reciente. Nos interesa efectuar algunas precisiones, tomando para la discusión en forma didáctica la información sobre el efecto de la estrategia invasiva sobre la mortalidad aguda respecto de la estrategia conservadora, y luego algunos comentarios generales. La pregunta médica para responder es si la estrategia invasiva ofrece ventajas en términos de mortalidad aguda, o si por el contrario puede implicar un riesgo, aun en las condiciones de haber sido efectuada en centros universitarios con medios óptimos como en los trabajos incluidos en el metaanálisis.

Primero algunas aclaraciones técnicas.

1. El uso en la Tabla 5 del riesgo relativo no fue un error tipográfico sino el intento de expresar en la tabla final un parámetro que es más intuitivo que el *odds*. Como se puede derivar de los números, los resultados no son idénticos a las tablas previas que sí se expresan en *odds ratio*. El error fue no haber aclarado por qué se había efectuado esta tabla con RR, o quizá aun más, no haber informado en el trabajo el parámetro para todos los cálculos.

2. Efectuados nuevamente los análisis en el STATA, los resultados obtenidos para mortalidad a 30 días sin FRISC riesgo relativo convencional 1,45 (1-2,12), OR con Mantel-Haenszel con Cornfield 1,47 (1-2,15) y OR con Mantel-Haenszel-Peto 1,46 (1-2,13), todos muy similares. Las discrepancias en la significación de la tabla de la carta del Dr. Borracci surgen no del hecho de utilizar el *odds ratio* o riesgo relativo, sino el criterio para el cálculo del intervalo de confianza. Se observa que en algún caso se informa que el intervalo de confianza va de 1 a 2,12, y con otra forma de calcular de 0,98 a 2,19. Más adelante discutimos lo que este resultado implica, pero debe quedar aclarado que no fue seleccionado un criterio favorable a una presunción previa, para poder cruzar la "línea

delgada" de la significación sino sólo informar los resultados tal cual surgen de las metodologías más habituales.

Ahora los temas conceptuales.

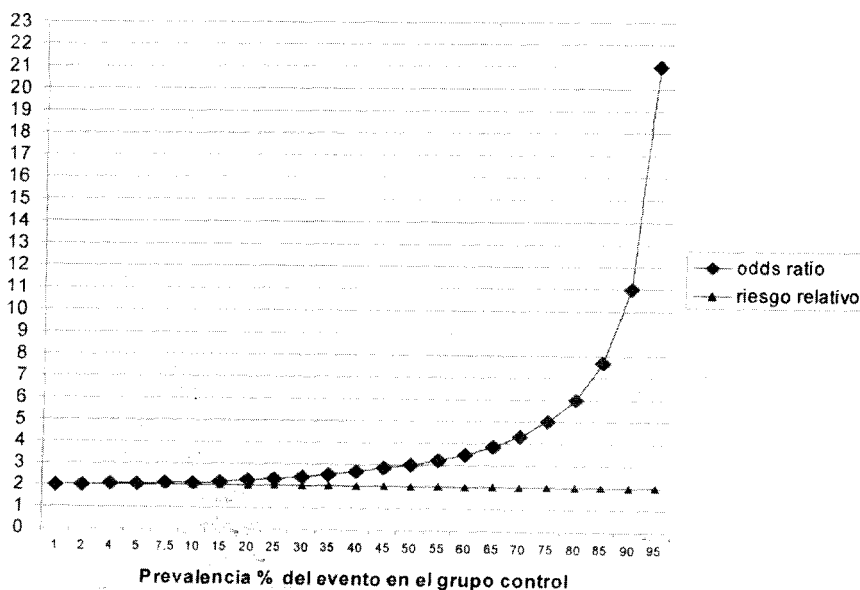
¿El odds ratio y el riesgo relativo pueden intercambiarse?

Esta pregunta es de gran interés, dado que ha existido a nuestro juicio un abuso del *odds ratio* quizá por un motivo fundamental: cuando se expresan los resultados, siempre el *odds ratio* exagera el beneficio o el perjuicio. Por ejemplo, si en un estudio fallecen 300 pacientes en 2.000 tratados con A, y 150 en 2.000 tratados con B, el riesgo relativo es: riesgo de muerte en el grupo B: $150/2.000 = 7,5\%$ dividido el riesgo de muerte en el grupo A $300/2.000 = 15\%$. Riesgo relativo = $7,5\%/15\% = 0,5$. El riesgo de morir del grupo B es de 0,5 respecto del grupo A, y por lo tanto la reducción de riesgo relativo fue del 50%. El *odds* correspondiente surge de la posibilidad relativa que se calcula como sigue: posibilidad de morir en el grupo B $150/1.850 = 8,1\%$ y el grupo A $300/1.700 = 17,6\%$. El *odds ratio* o posibilidad relativa es $8,1/17,6 = 0,46$, de manera tal que la reducción de la posibilidad de morir es del 54%, mayor que la reducción del riesgo de morir. El OR y el RR son muy similares cuando la prevalencia de los eventos es baja, y su discrepancia es cada vez mayor cuanto mayor es la prevalencia de los eventos. Como regla para el médico práctico, el *odds ratio* en estudio de intervención implica conceptual-

mente lo mismo que el riesgo relativo, exagerando un poco o mucho la tendencia con dependencia de la prevalencia del evento. Por ese motivo proponemos que el OR sea progresivamente menos usado en los estudios que evalúan intervenciones en grupos paralelos. En el análisis de la mortalidad aguda, como hemos explicado en el punto 2, el riesgo de morir aumenta un 45% (riesgo relativo 1,45) y la posibilidad de morir 46% o 47%, una diferencia clínicamente irrelevante. En la Figura resumimos la relación entre *odds* riesgo y riesgo relativo, que permite apreciar su dependencia acentuada de la prevalencia de los eventos que se analizan. En el gráfico de la carta del Dr. Borracci se presume una relación entre *odds ratio* y riesgo relativo que seguramente corresponde a una sola prevalencia de eventos, quizá 0,10 como adelanta en el texto, y no debe tomarse como una curva universal de relación *odds ratio* y riesgo relativo.

La ilusión de la p < 0,05 y la significación limítrofe

¿Se invalidan los resultados del trabajo si utilizamos la corrección del intervalo de confianza y con uno nos da significativo (p = 0,049) y con el otro no significativo (p = 0,052)? Este aspecto merece una interpretación clínica, que se expone en el punto siguiente, y una discusión estadística que intentaremos hacer entendible al lector. En primera instancia debemos recordar que el nivel por encima del cual consideraremos un resultado significativo o no, es arbitrario, y se establece *a priori* cuando se constru-



Relación entre *odds ratio* y riesgo relativo. Ejemplo con riesgo relativo de 2. En el gráfico se demuestra que el OR y RR son muy similares cuando la prevalencia del evento es baja, y muy diferentes cuando la prevalencia es elevada. En este ejemplo, el riesgo relativo se ha mantenido constante en 2, variando la prevalencia del evento en el grupo control del 1% al 95%. Con prevalencias del 1% (por ejemplo, 10 muertes en 1.000 pacientes en el grupo control contra 5 en 1.000 en el otro grupo) el *odds ratio* es 2,01 y el riesgo relativo es 2. Para una prevalencia del 50% (500 en 1.000 del grupo control contra 250 en 1.000 del otro grupo), el riesgo relativo continúa en 2 pero el *odds* ya es 3. Con prevalencias mayores la distancia entre el *odds* y el riesgo es muy notable.

ye un protocolo de investigación. Es frecuente que en estudios de importancia fundamental e inclusión de gran número de pacientes, se establezca que la significación será de 0,01, y en estudios más pequeños de 0,05. El creador de las pruebas de significación fue RA Fisher. Para Fisher, el umbral de $p = 0,05$ era mucho más flexible que el respeto casi religioso que le practicamos hoy, ya que consideraba que más que una interpretación definitiva de los resultados, los valores de p servían como una guía para la acción del investigador. Por ejemplo, afirmaba que un valor de p que estuviera alrededor de 0,05 no debía llevar al investigador ni a creer ni a descreer en la hipótesis de nulidad, sino más bien a la decisión práctica de realizar otro experimento. La " $p = 0,01$ " era, para él, el mínimo nivel deseable para publicar los resultados. Pero luego de publicado el experimento con una $p = 0,01$, esto debería llevar a pruebas posteriores con granjeros reales en las condiciones reales de una granja (se trataba de campos y granjeros, porque los métodos de aleatorización y las pruebas de significación surgieron a principios del siglo XX para conocer cuáles eran las mejores condiciones para mejorar la productividad de la tierra). En términos estadísticos, cuando se establece que la p que se tomará como criterio para significación será de 0,05, se considera un nivel de error de tipo I: cuando afirmamos que un resultado es significativo, sólo existe una posibilidad promedio de que este resultado haya sido sólo por azar (y de tal manera incurrir en un error al pensar que no fue por azar) en un estudio de cada 20 que llevemos a cabo. Podemos hacerlo más entendible si recordamos los criterios para el ingreso en la facultad. Se establece que ingresan los alumnos que sumen más de 60 puntos. Los que sacan 59, son muy similares en su capacidad de responder que los que suman 60, pero al establecer el punto de corte, 59 no ingresa y 60 sí. Esto es obviamente arbitrario y los pedagogos proponen cada vez más otras formas de evaluación. La misma arbitrariedad tiene la p . Es por este motivo que en los últimos años en la evaluación de ensayos clínicos se recomienda no informar que el estudio es significativo o no, en forma dicotómica, sino el valor exacto de p y fundamentalmente los intervalos de confianza del resultado, cuya importancia exponemos a continuación.

¿Cómo debe leer los resultados limítrofes un médico que intenta sacar conclusiones prácticas?

De acuerdo con lo que expusimos en el punto anterior, la p es sólo un elemento más del análisis conceptual y estadístico. El intervalo de confianza nos da una información más jugosa. Volvamos al ejemplo que estamos discutiendo y traduzcámoslo en términos conceptuales. La pregunta que se hace el médico es si la estrategia invasiva aporta un be-

neficio en términos de mortalidad aguda o puede implicar un daño. El metaanálisis muestra que no aporta un beneficio significativo, sino por el contrario el riesgo de morir aumenta un 45% como valor integrado del metaanálisis. ¿Este aumento es significativo o no? No parece tan relevante esta respuesta si lo pensamos en términos de intervalos de confianza. Tomemos de la tabla del Dr. Borraconi el riesgo relativo 1,45 (1-2,12) y expresémoslo en términos conceptuales: el riesgo de morir aumentó un 45% en forma significativa, y el riesgo real de morir puede ser similar o aumentar incluso 2 veces (100%) con las intervenciones. Tomando el segundo ejemplo, no significativo, OR 1,47 (0,98-2,19), leemos que la posibilidad de morir aumentó un 47% con el tratamiento invasivo. No podemos descartar que el tratamiento invasivo pueda incluso reducir la posibilidad de muerte un 2% (por el intervalo de confianza 0,98) pero tampoco que la posibilidad de morir aumente 2,19 veces, un 120%. *Debería ser obvio para el médico práctico que aplicando la estrategia invasiva no podemos pretender reducir la mortalidad, o a lo sumo un 2%, lo que no cambia mucho dado que tenemos una fuerte posibilidad de incrementarla, promedio 47% y hasta más del doble.* Es éste el motivo de la recomendación de no informar en los *abstracts* "la diferencia fue significativa o no significativa" sino el nivel de p y el intervalo de confianza que permite una inferencia clínica mucho mejor. De estos conceptos surge el cauteloso comentario para interpretar los resultados del metaanálisis, que el colega resalta en sus párrafos finales.

¿Son iguales los criterios para interpretar un daño que un beneficio?

De aun mayor importancia es destacar que la tendencia del metaanálisis marca un daño en agudo, y que en la lectura médica habitual *aceptar que algo puede ser peligroso no requiere estrictamente significación estadística, sino sólo una tendencia perjudicial asociada con una probabilidad muy baja de que sea beneficioso.* No se diseñan estudios para demostrar daño, y cuando en un ensayo de investigación con una nueva droga se producen eventos desfavorables graves, como tendencia inicial los comités de ética detienen la investigación, sin esperar significación estadística. Como ejemplo de la práctica clínica, en las guías y libros de evidencia se recomienda evitar los bloqueantes cálcicos y la lidocaína en forma sistemática en la fase aguda del infarto. El metaanálisis de bloqueantes cálcicos (incremento del 4% de la mortalidad no significativo) y del uso sistemático de lidocaína (OR 1,38, IC 95% 0,98 a 1,95) no muestran un perjuicio significativo, sino una tendencia a incrementar la mortalidad, pero con un intervalo de confianza que hace muy poco probable que sean

beneficiosas y bastante probable que sean perjudiciales, lo que ha llevado a la comunidad médica a evitar estas drogas. *La traslación de la información médica de los ensayos a la acción no depende de la p sino de un complejo juicio clínico.* La sospecha de un daño con p 0,06 o 0,1 también puede ser un criterio para evitar esa conducta, por supuesto si se ha evaluado en forma adecuada en un número suficiente de pacientes.

Es muy estimulante que la Revista Argentina de Cardiología incluya la posibilidad de polemizar intelectualmente con el afán de mejorar nuestra comprensión e interpretación de los problemas y que nuestro trabajo haya merecido comentarios del nivel del enviado por el Dr. Borracci.

Dres. Carlos Tajer, Hernán Doval, José Tessler