

# Ausencia de evidencia y evidencia de ausencia de asociación. No es lo mismo. Parte 10

Título en inglés: Absence of evidence or evidence of absence. That is the question.

Bottaro F

Jefe de Emergencias Hospital Británico  
Servicio de Clínica Médica Hospital Británico  
Coordinador Comité de Revisión Institucional Hospital Británico

febottaro@hotmail.com

Fecha de recepción: 1/04/2017  
Fecha de aprobación: 17/04/2017



CÓMO SE LEE

HEMATOLOGÍA  
Volumen 21 n° 1: 96-99  
Enero - Abril 2017

**Palabras claves:** Ausencia de evidencia, Intervalo de confianza.

**Keywords:** Absence of evidence, Confidence interval.

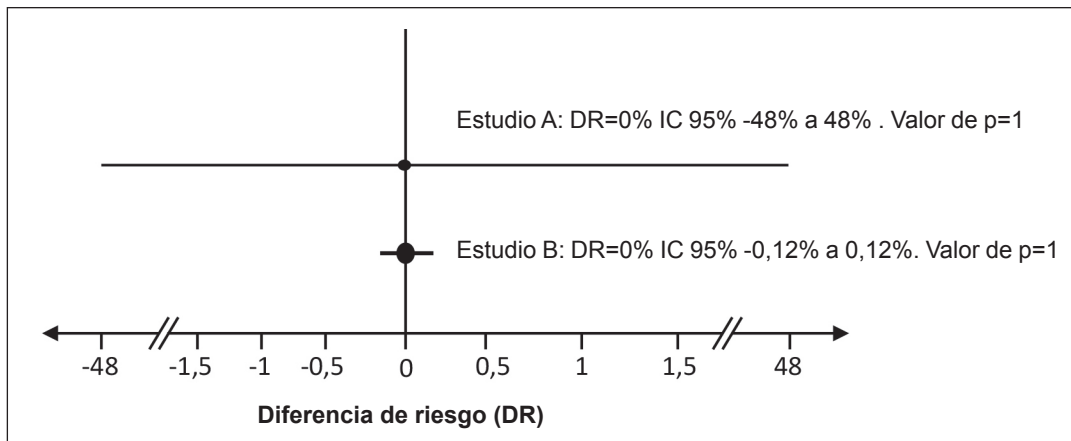
*“Nada es veneno, todo es veneno. La diferencia está en la dosis”  
Paracelso*

A menudo en la interpretación de ensayos clínicos encontramos con resultados que poseen significancia estadística (un valor de  $p < 0,05$ ) pero nos es muy difícil establecer su relevancia clínica. La diferencia entre “significancia estadística” e “importancia clínica” es reconocida desde hace tiempo, pero aún hoy es un error de interpretación común. Con frecuencia un resultado estadísticamente significativo en un ensayo que reclutó gran tamaño de muestra no es clínicamente importante, pero una situación más compleja surge de la interpretación “errónea” de un hallazgo no significativo. Por convención, un valor de  $p$  mayor a 0,05 ( $p > 0,05$ ) se denomina “no significativo”. Los ensayos clínicos randomizados

controlados que no muestran diferencias significativas en el efecto de una intervención entre sus grupos son frecuentemente denominados estudios “**negativos**”. Esta terminología implica que el estudio ha demostrado que no hay diferencias, mientras que frecuentemente lo que estos estudios demuestran es “ausencia de evidencia de una diferencia”. Éstas son dos situaciones claras y diferentes.

### Un ejemplo

Consideremos dos ensayos randomizados controlados (ERC) con drogas hipotéticas controlados con placebo: el estudio A posee 10 pacientes en cada rama del estudio, mientras que el estudio B posee 5000 pacientes en cada rama (ver **Figura 1**).



**Figura 1.** En esta figura se representan las diferencias de riesgo del ejemplo del texto. Cuando el intervalo de confianza (IC 95%) de la diferencia de riesgo sobrepasa la línea del 0 significa que no hay efecto de la intervención. Se puede observar cómo el gráfico del estudio A involucra un IC mucho más amplio que el estudio B. Cualquier modificación de la DR dentro de su IC no alteraría la interpretación del estudio, siendo sus hallazgos indicadores de evidencia de ausencia de beneficio. En el gráfico del estudio A modificaciones leves en el tamaño, en el número o distribución de los eventos pueden alterar radicalmente la DR observada.

En ambos estudios 5 pacientes murieron en cada rama. ¿Cuál es el significado de esta situación? ¿Se podría decir que ambos estudios poseen el mismo valor y de ambos podemos concluir que la droga estudiada en cada estudio es inefectiva? En una primera lectura podríamos decir que sí, pero en una interpretación más profunda podemos notar que el estudio A no tiene el mismo peso que el estudio B para realizar dicha afirmación. De hecho, si modificásemos levemente el tamaño del estudio A o la distribución de eventos, sus resultados y, en consecuencia, su interpretación se verían seriamente amenazados. Un aumento en el estudio A en 200 personas por rama y una diferencia leve en la distribución de muertes, 40 muertes en grupo tratamiento y 60 muertes en grupo control, ya logra significación estadística ( $p=0,03$ ) y relevancia clínica significativa (número necesario para tratar (NNT) igual a 10). Por otro lado, en el estudio B, aun aumentando su tamaño en 200 pacientes por grupo y modificando la distribución de eventos al igual que el estudio A (40 muertes en grupo tratamiento y 60 muertes en grupo control), la significancia estadística no se modifica ( $p=0,06$ ) y tampoco su relevancia o importancia clínica.

En este ejemplo podemos ver que el estudio A es más sensible a modificaciones leves tanto de su tamaño como a la posible distribución de eventos entre sus ramas, mientras que el estudio B es más robusto y su interpretación clínica y estadística no se modifica

aun ante cambios de igual magnitud que los del estudio A. En otras palabras, el estudio “A” es un caso de **“ausencia de evidencia de beneficio”**, mientras que el estudio “B” es un caso de **“evidencia de ausencia de beneficio”**.

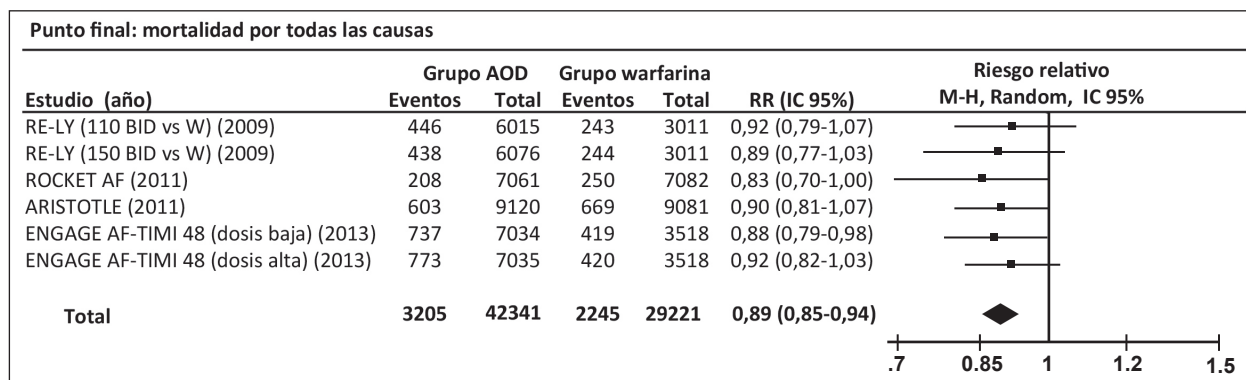
#### Semejanzas y diferencias: valor de p e intervalos de confianza

En ambas situaciones del ejemplo del párrafo anterior los valores de p serán cercanos al 1,0. Esto nos da la pauta que el valor de p no es capaz de distinguir entre estas dos situaciones: **“ausencia de evidencia de beneficio”** y **“evidencia de ausencia de beneficio”**. Sin embargo, una mirada precisa al intervalo de confianza nos puede ayudar a distinguirlos. El intervalo de confianza del 95% (IC 95%) de la diferencia de riesgo del estudio A es -48% a 48%, sugiriendo que el estudio es no concluyente y que los datos son consistentes con una potencial disminución de mortalidad de hasta 48% (beneficio) o inclusive un posible aumento de mortalidad de hasta 48% (daño). El IC 95% del estudio B oscila entre -0,12% y 0,12% (ambos valores muy pequeños e irrelevantes), sugiriendo que el tratamiento es inefectivo, no pudiendo establecer beneficio o daño del mismo, pero con certeza que su efecto será irrelevante.

Este ejemplo demuestra una limitación importante del valor de p en la contribución a la interpretación

de ERC, y cómo los intervalos de confianza pueden colaborar en el proceso. Un ejemplo cercano sobre ausencia de evidencia de beneficio se puede observar en el estudio Re-Ly<sup>(1)</sup>. Éste fue un estudio que reclutó 18113 pacientes y fue diseñado para comparar dos dosis fijas de dabigatrán con la administración de warfarina para la prevención de accidente cerebrovascular y embolia sistémica en sujetos con

fibrilación auricular. En el análisis del punto final mortalidad, dabigatrán 150 mg cada 12 horas (BID) no pudo demostrar beneficio en la reducción: riesgo relativo (RR) 0,88 (IC 95% 0,77-1,0) valor de p=0,051. Sin embargo, y llamativamente, el tamaño del efecto fue similar al observado en otros estudios de anticoagulantes directos comparados con warfarina (ver **Figura 2**)<sup>(2, 3, 4)</sup>.



**Figura 2.** Meta-análisis de los 4 estudios que compararon anticoagulantes directos versus warfarina para la prevención de accidente cerebro-vascular en pacientes con fibrilación auricular.  
W: warfarina, M-H: Mantel and Haentzel, AOD: anticoagulantes directos, BID: cada 12 horas.

De hecho, al meta-analizar los resultados en mortalidad por todas las causas de todas las drogas anticoagulantes directas comparadas con warfarina podemos observar una reducción del punto final coherente y consistente entre todas ellas y el estimado del efecto global se hace estadísticamente significativo y, obviamente, clínicamente relevante (RR 0,89 IC 95% 0,85-0,94), hallando una reducción de mortalidad potencial de hasta 15%. ¿Qué ocurrió al analizar los resultados del estudio Re-Ly? Al dividir la población total del estudio (n=18113 personas) en tres grupos de cerca 6000 personas cada uno, el estudio perdió poder estadístico para mostrar una diferencia en el punto final mortalidad.

Un peligro similar en la interpretación de ERC surge del análisis de innumerables estudios “negativos” (no significativos) que compararon estrategias de reperfusión con fibrinolíticos en el tratamiento de infarto agudo de miocardio con elevación de ST versus placebo. El meta-análisis de estos estudios arroja un efecto modesto pero clínicamente significativo (y estadísticamente altamente significativo) en reducción de mortalidad de 22%, sin embargo sólo 5 de 24 ERC habían mostrado un efecto significativo con un valor de p<0,05<sup>(5)</sup>.

Otra limitación del valor de p en la interpretación de

ensayos clínicos la podemos observar en el siguiente ejemplo: consideremos ahora dos ERC hipotéticos que compararon tratamientos versus placebo: el estudio A posee 200 personas en cada una de sus ramas, el estudio B 2000 sujetos por rama. En el estudio A, 40 pacientes (20%) murieron en el grupo tratamiento y 60 (30%) en el grupo placebo: arrojando una diferencia de riesgo de -10% (valor de p=0,03). En el estudio B, 38 personas fallecieron en el grupo tratamiento (1,9%) y 60 personas en el grupo placebo (3%), observándose una diferencia de riesgo de -1,1% (valor de p=0,03). Claramente se puede observar que los resultados de ambos estudios no son iguales, sin embargo su valor de p es similar. Éste es otro claro ejemplo de las limitaciones del valor de p. Mientras que usualmente es razonable no aceptar un nuevo tratamiento hasta que haya evidencia positiva a su favor, cuando hay involucrado temas de salud pública mayores debemos cuestionarnos sobre si la ausencia de evidencia es justificación válida suficiente para la inacción.

Cuando a menudo nos encontramos con la afirmación de que no “no hay evidencia de que A cause B”, nos debemos cuestionar en primera instancia si la ausencia de evidencia implica simplemente que carecemos de información suficiente para aseverar

o refutar esa afirmación. Si hay datos debemos entonces buscar la cuantificación del efecto de la asociación más que solamente el valor de p. Cuando los riesgos son menores, los valores de p nos pueden conducir a una interpretación errónea, y los IC serán muy amplios, implicando mucha incertidumbre.

Estos dos escenarios estarán siempre sujetos a nuestro análisis en la interpretación de ensayos clínicos.

### **Bibliografía**

1. Connolly SJ, Ezekowitz MD, Yusuf S y col; RE-LY Steering Committee and Investigators. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med.* 2009;361(12):1139-51.
2. Patel MR, Mahaffey KW, Garg J y col; ROCKET AF Investigators. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med.* 2011;365(10):883-91.
3. Granger CB, Alexander JH, McMurray JJ y col; ARISTOTLE Committees and Investigators. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med.* 2011;365(11):981-92.
4. Giugliano RP, Ruff CT, Braunwald E y col; ENGAGE AF-TIMI 48 Investigators. Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med.* 2013;369(22):2093-104.
5. Yusuf S, Collins R, Peto R y col. Intravenous and intracoronary fibrinolytic therapy in acute myocardial infarction: overview of results on mortality, reinfarction and side-effects from 33 randomized controlled trials. *Eur Heart J.* 1985 Jul;6(7):556-85.