



El tamaño de efecto. Esto funciona, pero ¿cuánto?

Effect size. This works but, how much?

Javier Pascual Huerta

Clinica del Pie Elcano, Bilbao, España

En las dos cartas anteriores de esta sección se revisaron los intervalos de confianza como método para permitir realizar inferencias de los datos obtenidos en una investigación^{1,2}. Una ventaja que aportan los intervalos de confianza es que nos ayudan a interpretar el tamaño de efecto. Pero, ¿qué es el tamaño de efecto? El tamaño de efecto es un concepto muy simple. Responde a la pregunta *¿cuánto de eficaz es este tratamiento?* o *¿cuánta es la diferencia que existe entre estos dos grupos?* Es la respuesta cuantitativa a la pregunta de investigación.

Veamos un ejemplo; Malay y cols.³ realizaron un ensayo clínico aleatorizado para comprobar el efecto de las ondas choque como tratamiento en pacientes con fasciopatía plantar. Para ello estudiaron a 172 pacientes con fasciopatía plantar que dividieron en dos grupos de forma aleatoria con una distribución de 2:1 de la siguiente forma: 115 pacientes fueron tratados mediante una única sesión de ondas de choque y 57 fueron sometidos a un tratamiento placebo

simulando la sesión de ondas de choque pero sin realmente recibirlas. Los resultados mostraron una reducción del dolor en el grupo de tratamiento con ondas de choque en comparación con el grupo de tratamiento con placebo a los dos meses ($p = 0.026$) y a los tres meses ($p = 0.045$). Los autores tomaron estos valores p como significativos al estar por debajo de 0.05, y concluyeron con esos resultados que las ondas de choque funcionan y son eficaces para el tratamiento de la fasciopatía plantar en comparación con un tratamiento placebo.

La pregunta inmediata que surge a continuación es: "Muy bien; las ondas de choque son eficaces... pero ¿cuánto?". Por el estudio sabemos que reducen el dolor, pero la pregunta es *¿cuánto es la reducción del dolor que ofrecen las ondas de choque?* Esto es lo que se conoce como el tamaño de efecto. Malay y cols. midieron en su estudio la reducción del dolor mediante una escala analógica visual (EAV) como herramienta para valorar la eficacia de los tratamientos que vienen

Tabla I. Cambios en el dolor de talón valorado por el examinador.

	Grupo ondas choque	Grupo placebo	Valor p
MES 2			
Reducción dolor (EAV)	-2.30	-1.31	0.026
Diferencia (IC 95 %)	-0.99 (-1.86,-0.12)		
MES 3			
Reducción dolor EAV)	-2.51	-1.57	0.045
Diferencia (IC 95 %)	-0.94 (-1.87,-0.02)		

Adaptado de Malay y cols³.

EAV: escala analógica visual. IC 95 %: intervalo de confianza al 95 %.



Tabla II. Tamaño de efecto.

El tamaño de efecto es el valor que muestra cuánto es la diferencia entre dos grupos de estudio. Es la respuesta cuantitativa a la pregunta de investigación: ¿cuánto?

Lo más práctico es presentar el tamaño de efecto en las unidades de la variable estudiada en la investigación, aunque también se pueden utilizar medidas de estandarización (ej: *d* de Cohen)

Siempre debe presentarse con su intervalo de confianza que expresa el nivel de incertidumbre existente al generalizar los resultados de la muestra a toda la población

resumidos en la Tabla I. En el grupo de tratamiento con ondas de choque existió una reducción del dolor en la EAV de 2.3 puntos a los dos meses y de 2.51 puntos a los tres meses, mientras que en el grupo de tratamiento con placebo existió una reducción del dolor en EAV de 1.31 a los dos meses y de 1.57 a los tres meses. El tamaño de efecto es la diferencia en la reducción del dolor entre el grupo de tratamiento con ondas de choque y el grupo de tratamiento con placebo. Esta diferencia fue de 0.99 a los dos meses con un intervalo de confianza al 95 % (IC 95 %) de 1.86 a 0.12 y de 0.94 a los tres meses con un IC 95 % de 1.87 a 0.02 a los tres meses.

El tamaño de efecto responde a la pregunta de cuánto de eficaz es el tratamiento o cuánta es la diferencia entre los grupos del estudio. Provee una respuesta cuantitativa a la pregunta de investigación. Cualquier valor *p* lleva asociado un tamaño de efecto con su correspondiente intervalo de confianza. Lo ideal es siempre presentar el tamaño de efecto en las mismas unidades de la variable que se está estudiando (pueden ser los ángulos de deformidad en una radiografía, los grados de movimiento de una articulación o la reducción del dolor mediante una EAV, tal y como se ha presentado en el estudio de Malay y cols.). Los lectores del estudio se suelen encontrar más familiarizados con esas unidades y pueden juzgar mejor la significación práctica de los resultados. Sin embargo, en muchas ocasiones no es posible presentar el tamaño de efecto en esas unidades porque usan diferentes escalas o su significación clínica de los resultados no ayuda a entender el estudio porque los lectores o investigadores no son expertos en esas escalas (ej: cambio en Kilopascales de presión con una plantilla, cambio de los momentos articulares expresados en newtons*metro/kilogramo o el cambio del valor tangencial a una pendiente de cualquier representación gráfica). En estos casos, para el lector puede ser más útil estandarizar el tamaño de efecto presentándolos en unidades estandarizadas. Existen diversos métodos de realizar eso, pero el más comúnmente utilizado es el valor *d* de Cohen. La *d* de Cohen expresa la diferencia entre dos grupos en unidades de desviaciones estándar. El valor es $d = (\text{media grupo 1} - \text{media grupo 2}) / \text{desviación estándar de ambos grupos}$. Se considera que el tamaño de efecto es bajo cuando la *d* de Cohen es de 0.2, medio cuando es de 0.5, alto cuando es

de 0.8 y muy alto a partir de 1.3⁴. En este caso concreto, el valor de *d* de Cohen en el estudio de Malay y cols. podría ser de 0.36, lo cual ejemplifica un valor bajo-medio de tamaño de efecto¹.

¿Por qué es importante el tamaño de efecto? Lo primero y más evidente es que el tamaño de efecto aporta un significado práctico a los resultados de una investigación. Un resultado estadísticamente significativo no siempre significa que en la práctica clínica sea importante; de hecho, puede ser clínicamente ridículo o insignificante. El estudio de Malay y cols.³ muestra que la reducción del dolor con ondas de choque en comparación con el placebo fue de aproximadamente 1 punto en una EAV de 10 puntos. Esta es una información muy relevante que nos ayuda a poner la información del estudio en un contexto práctico más útil y nos hace valorar mejor los resultados del estudio. El valor *p* nos dice únicamente si hay una diferencia o no la hay, pero no dice si esa diferencia es grande o pequeña.

El segundo punto importante es que el tamaño de efecto junto con su intervalo de confianza expresa la incertidumbre que existe al generalizar los resultados de este estudio a toda la población. El tamaño de efecto es una estimación del efecto encontrado en este estudio concreto con esa muestra concreta. El intervalo de confianza reportado por Malay y cols. en la reducción del dolor a los tres meses va de 1.87 a 0.02. Este intervalo de confianza expresa la incertidumbre existente al generalizar los resultados de esta muestra a toda la población. En este caso, el intervalo de confianza del tamaño de efecto es relativamente largo y refleja cierto grado de incertidumbre. De hecho, no sería raro que el verdadero efecto de las ondas de choque sobre la reducción del dolor fuera una reducción notable de casi 2 puntos en una escala analógica visual de 10 puntos, pero tampoco sería raro que el efecto fuera prácticamente nulo o inexistente (0.02) en la reducción del dolor.

* NOTA: El estudio de Malay y cols. no aporta datos sobre la desviación estándar, por lo que el valor *d* de Cohen ha sido calculado por medio de una aproximación a los datos de desviación estándar en base a un estudio similar realizado por Kudo y cols. (Kudo P, Dainty K, Clarfield M, Coughlin L, Lavoie P, Lebrun C. Randomized, Randomized, Placebo-Controlled, Double-Blind Clinical Trial Evaluating the Treatment of Plantar Fasciitis with an Extracorporeal Shockwave Therapy (ESWT) Device: A North American Confirmatory Study. J Orthop Res. 2006;24(2):115-23. DOI: 10.1002/jor.20008) por lo que este valor no puede considerarse fiable.

Finalmente, el tercer punto relevante del tamaño de efecto aquí es la presencia de una muestra adecuada o no. Existe preocupación sobre la investigación en pie y tobillo como históricamente de baja potencia. ¿Qué significa de baja potencia? Un estudio de baja potencia es aquel en el que la muestra es demasiado pequeña como para aportar estimaciones del tamaño de efecto fiables y exactas. El resultado es que los tamaños de efecto presentan una horquilla tan extensa que es difícil sacar conclusiones con los mismos. Aportar el tamaño de efecto con su intervalo de confianza señala la presencia o no de una muestra adecuada en el estudio: intervalos de confianza muy largos son indicativos de una muestra inadecuada. Por lo tanto, el hecho de reportar los tamaños de efecto con sus intervalos de confianza ayuda también a mejorar la atención prestada a la potencia estadística y aumenta la fiabilidad de la literatura.

En definitiva, detrás de cada test de significación de la hipótesis nula existe un tamaño de efecto con su intervalo de confianza que necesita ser especificado e interpretado. Es importante que los investigadores reporten el tamaño de efecto con su intervalo de confianza en sus artículos o alguna otra forma de expresar la incertidumbre al generalizar los resultados obtenidos de su muestra sobre la población (Tabla II).

CONFLICTO DE INTERESES

El autor no presenta ningún conflicto de intereses relevante con la presente carta.

FINANCIACIÓN

No ha existido ninguna fuente de financiación alguna en la presente carta

BIBLIOGRAFÍA:

1. Pascual Huerta J. Intervalos de Confianza vs. valor p. Porque el gris es importante. *Rev Esp Podol.* 2018;29(1):60-2. DOI: 10.20986/revspod.2018.1517/2018
2. Pascual Huerta J. Intervalos de confianza vs. valor p (II). Pensamiento meta-analítico vs. pensamiento dicotómico. *Rev Esp Podol.* 2018;29(2):113-6. DOI: 10.20986/revspod.2018.1522/2018
3. Malay DS, Pressman MM, Assili A, Kline JT, York S, Buren B, et al. Extracorporeal Shockwave Therapy Versus Placebo for the Treatment of Chronic Proximal Plantar Fasciitis: Results of a Randomized, Placebo-Controlled, Double-Blinded, Multicenter Intervention Trial. *J Foot Ankle Surg.* 2006;45(4):196-210. DOI: 10.1053/j.jfas.2006.04.007.
4. Sullivan GM, Feinn R. Using Effect Size – or why the P value is not enough. *J Grad Med Educ.* 2012;4(3):279-82. DOI: 10.4300/JGME-D-12-00156.1