



CIRUGÍA ESPAÑOLA

www.elsevier.es/cirugia



Carta metodológica

El tamaño del efecto. ¿la magnitud importa?

The size of the effect. Does magnitude matter?



Antonio Ríos^{a*} y Marina Iniesta-Sepúlveda^b

^a Servicio de Cirugía General y del Aparato Digestivo, IMIB - Hospital Clínico Universitario Virgen de la Arrixaca, Departamento de Cirugía, Pediatría, Obstetricia y Ginecología, Universidad de Murcia, Murcia, España

^b Departamento de Psicología, Facultad de Ciencias de la Salud, Universidad Católica de Murcia (UCAM), Murcia, España

La evaluación de los resultados de la investigación se ha apoyado generalmente en las pruebas de significación estadística¹, en las que el investigador concluye que un efecto debe existir en la población, si la probabilidad de haberlo obtenido por azar en la muestra es muy reducida (generalmente menor al 1% o al 5%). Sin embargo, no sólo es importante saber de la existencia del efecto en la población a estudio, sino conocer la magnitud que tiene dicho efecto. En este sentido, hay que recordar que la prueba de significación estadística (valor p) no aporta información acerca de la magnitud de un efecto y no va asociada necesariamente a la significación clínica o práctica de los resultados. La significación clínica hace referencia al impacto real que tiene el efecto observado sobre el paciente². Desafortunadamente, es frecuente encontrar interpretaciones que confunden significación estadística con significación práctica.

Actualmente este aspecto ha adquirido especial relevancia. La significación estadística es altamente dependiente del tamaño muestral, hoy en día los estudios en cirugía y otras disciplinas médicas, son en su mayoría multicéntricos y cuentan con amplias muestras de participantes. Este hecho, puede llevar a la obtención de resultados significativos en la mayor parte de las ocasiones y que no siempre son relevantes a nivel práctico³.

Para conocer la significación práctica de los resultados de la investigación, se desarrollaron los índices del tamaño de efecto, estos informan de la medida en la que ocurre el fenómeno objeto de estudio en la población⁴. Existen numerosos índices del tamaño del efecto y generalmente

son clasificados según hagan referencia a la magnitud de las diferencias entre grupos o al grado de asociación entre variables⁴ (tabla 1).

Índices para la valoración de la magnitud de las diferencias entre grupos

Los más utilizados son los que se encuentran dentro de la familia *d*, los cuales están basados en las diferencias entre medias. Entre ellos, la *diferencia media tipificada* es el más utilizado y permite comparar dos grupos en una sola medición del resultado. La fórmula de Cohen⁵ está indicada cuando ambos grupos presentan tamaño y variabilidad similares. Existen otros índices dentro de esta familia, por ejemplo, en el caso de diseños con pretest y posttest es más adecuado utilizar índices basados en las puntuaciones de cambio⁶, como el *índice de cambio medio tipificado* (un solo grupo) o la *diferencia entre los cambios medios tipificados* (dos grupos).

Cuando se pretende comparar dos grupos en una variable dicotómica se utilizan índices basados en las proporciones de riesgo, es decir, en la probabilidad de que ocurra un evento de interés en función de la presencia o ausencia de un factor. Los índices más utilizados son la *razón de riesgos* o *riesgo relativo* (RR) y la *razón de ventajas* u *odds ratio* (OR). El índice RR indica en qué medida la probabilidad de que ocurra el evento en un grupo, es superior a la probabilidad de que ocurra el evento en el otro, mientras que el OR debe ser interpretado en términos de ventajas y no de probabilidad⁷. La ventaja nos indica cuántas

* Autor para correspondencia.

Correos electrónicos: arriors@um.es, arriors4@gmail.com (A. Ríos).

<https://doi.org/10.1016/j.ciresp.2021.10.015>

0009-739X/© 2021 AEC. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Tabla 1 – Prueba estadística asociada e interpretación de algunos de los principales índices del tamaño del efecto

Índices del tamaño del efecto para la valoración de las diferencias entre grupos		
Índice	Prueba estadística	Interpretación
Diferencia media tipificada	t de Student para muestras independientes	Magnitud de la diferencia de medias entre dos grupos. Criterio de Cohen (1988): en torno a 0,20 baja en torno a 0,50 moderada > 0,8 alta
Riesgo Relativo (RR)	Tabla de contingencia 2X2	Magnitud de la diferencia entre las proporciones de riesgo en dos grupos.
Odds Ratio (OR)	Tabla de contingencia 2X2	Magnitud de la diferencia entre las ventajas de dos grupos.
Índices del tamaño del efecto para la valoración de la relación entre variables		
Índice	Prueba estadística	Interpretación
Correlación de Pearson (r)	Análisis de correlación lineal	Grado de asociación lineal entre dos variables. Criterio de Cohen (1988): < 0,10 asociación nula 0,10 - 0,29 baja 0,30 - 0,49 moderada >0,50 elevada
Coefficiente de Determinación (R ²)	Regresión lineal simple	Proporción de varianza de la variable dependiente explicada por la variable independiente
Coefficiente de Determinación Ajustado (R ² adj)	Regresión lineal múltiple	Proporción de varianza de la variable dependiente explicada por el modelo
Eta-Cuadrado (η^2)	ANOVA de un factor	Proporción de varianza de la variable dependiente explicada por el factor.
Eta-Cuadrado Parcial (η^2_p)	ANOVA factorial	Proporción de varianza de la variable dependiente explicada por cada factor
Omega-Cuadrado (ω^2)	ANOVA de un factor ANOVA factorial	Proporción de varianza de la variable dependiente explicada por el modelo.

veces la probabilidad de que el evento ocurra, es superior a la probabilidad de que no ocurra, o viceversa, así el OR sería igual a la razón entre las ventajas de ambos grupos. Los valores de RR y OR serán similares cuando las proporciones de riesgo sean bajas.

Índices para la valoración del grado de asociación entre variables

Dentro de los índices de asociación, el más conocido es la *correlación de Pearson*, el cual mide la magnitud y el sentido de la relación lineal entre dos variables. Este índice varía entre -1 y 1, siendo menor la fuerza de la asociación entre las variables en la medida que la correlación se aproxime a 0. La alternativa no paramétrica a la correlación de Pearson (también utilizada cuando las variables son ordinales) es la *correlación de Spearman* (Rho).

En el contexto del análisis de asociación entre variables, es interesante conocer qué proporción de varianza de la variable dependiente, viene explicada por la variable independiente. En los modelos de regresión lineal, esta proporción viene dada por el *coeficiente de determinación* (R²) que es igual al cuadrado del coeficiente de correlación de Pearson. Cuando se incluyen varios predictores, se prefiere el *coeficiente de determinación ajustado* (R² a dj) que presenta un mayor control de la varianza de error teniendo en cuenta el tamaño muestral y el número de predictores del modelo⁸. Un caso especial es la regresión logística, donde la variable dependiente es dicotómica, cuyo tamaño del efecto estimado es el logaritmo natural del OR⁷.

Otro índice de varianza explicada, derivado en este caso del análisis de varianza (ANOVA), es el índice *eta-cuadrado* (η^2). Este informa de la proporción de varianza explicada de la variable dependiente por una variable independiente categórica. Cuando se trata de un ANOVA factorial, η^2 *parcial* indica el porcentaje de varianza explicada por cada uno de los predictores. Existen otros índices del tamaño del efecto para el ANOVA, como el *omega-cuadrado* (ω^2), el cual corrige la posible sobreestimación de la varianza explicada de η^2 ⁹.

Consideraciones para la publicación

Actualmente, en las principales guías para la publicación de la investigación médica se indica que además de la significación estadística, debe considerarse informar del tamaño de efecto, ya que este ofrece una medida de la importancia clínica de los resultados². Entre las principales razones para informar del tamaño del efecto encontramos^{1,4,8}: 1) Conocer la significación práctica de los resultados; 2) Posibilitar el cálculo de la potencia estadística. Así, en el diseño de un nuevo estudio, la magnitud del efecto observada en la literatura científica previa es utilizada para calcular la potencia estadística y estimar el tamaño muestral necesario y 3) Permitir la comparación entre estudios y la integración de la evidencia empírica en el metaanálisis.

A la hora de informar del tamaño del efecto deben tenerse en cuenta una serie de consideraciones¹⁰. En primer lugar, debe especificarse cuál de los índices se ha utilizado para obtener el tamaño del efecto y este debe ser adecuado al tipo

de análisis realizado. En segundo lugar, el tamaño del efecto describe las propiedades de una muestra, siendo un estimador potencial de su correspondiente parámetro en la población; es por ello, que debe aportarse el intervalo de confianza, el cual informa sobre el grado de precisión de la estimación. En tercer lugar, cabe recordar, que a pesar de contar con criterios matemáticos para la interpretación del tamaño del efecto existen otra serie de factores a la hora de determinar la relevancia práctica de los resultados. Así, los investigadores conocedores del contexto y del fenómeno objeto de estudio, deben aportar una explicación del significado del efecto encontrado en el mundo real⁵. Un efecto pequeño (según criterios matemáticos) con grandes consecuencias a nivel sanitario o económico puede ser relevante para la sociedad.

A modo de conclusión, cabe señalar que la evaluación objetiva de la evidencia científica requiere de la valoración complementaria de las pruebas de significación estadística y de las medidas de la magnitud del efecto, la cual, unida a la interpretación sustantiva y contextualizada por parte de los investigadores, permitirá obtener una idea más certera del significado de los hallazgos para la realidad clínica.

BIBLIOGRAFÍA

1. Kyriacou DN. The Enduring Evolution of the P Value. *JAMA*. 2016;315:1113-5. <http://dx.doi.org/10.1001/jama.2016.2152>.
2. American Medical Association. *AMA manual of style: A guide for authors and editors*. New York: Oxford University Press. 2020.
3. Livingston EH, Elliot A, Hynan L, Cao J. Effect size estimation: a necessary component of statistical analysis. *Arch Surg*. 2009;144:706-12. <http://dx.doi.org/10.1001/archsurg.2009.1504>.
4. Ellis PD. *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. New York: Cambridge University Press. 2010.
5. Cohen J. *Statistical power analysis for the behavioral sciences*, 2nd ed. New Jersey: Lawrence Erlbaum Associates. 1988.
6. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley. 2019.
7. Norton EC, Dowd BE, Maciejewski ML. Odds Ratios-Current Best Practice and Use. *JAMA*. 2018;320:84-5. <http://dx.doi.org/10.1001/jama.2018.6971>.
8. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol Gen*. 2012;141:2-18. <http://dx.doi.org/10.1037/a0024338>.
9. Okada K. Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. *Behaviormetrika*. 2013;40:129-47. <http://dx.doi.org/10.2333/bhmk.40.129>.
10. Karadaghy OA, Hong H, Scott-Wittenborn N, Sinha P, Suko J, Tait S, et al. Reporting of effect size and confidence intervals in JAMA Otolaryngol Head Neck Surg. *JAMA Otolaryngol Head Neck Surg*. 2017;143:1075-80. <http://dx.doi.org/10.1001/jamaoto.2017.1504>.