

Magnitud del efecto para la diferencia de dos grupos en ciencias de la salud

Effect size for the difference of two groups in health sciences

T. Caycho¹, J. Ventura-León¹, R. Castillo-Blanco²

Sr. Director:

Recientemente López Cortés y col¹ ofrecieron una importante contribución al estudiar la calidad de vida asociada con la salud en un grupo de pacientes diagnosticados de enfermedad de Crohn y colitis ulcerosa. Los autores aportan diferencias estadísticamente significativas en función del tipo de enfermedad ($p=0,005$) y el sexo ($p=0,001$), en base al test de significación de la hipótesis nula (NHST, por sus siglas en inglés). En la NHST, la hipótesis nula (H_0), referida a la ausencia de diferencias estadísticamente significativas, se rechaza o mantiene en función de la probabilidad (p -valor) vinculada a la prueba estadística utilizada². Diversas revisiones^{2,3} critican el empleo del NHST, señalando, entre sus limitaciones, que solo hace referencia al máximo error permitido para rechazar la H_0 , así como su sensibilidad para mantener o rechazar la H_0 , en función al tamaño de la muestra, no afirmando si la diferencia entre grupos es importante. Así, el empleo del NHST es un procedimiento que podría ser considerado como adecuado, pero no suficiente, siendo necesario conocer el tamaño del efecto para valorar la importancia práctica de las diferencias entre los grupos de comparación⁴.

El tamaño del efecto se calcula mediante la d de Cohen, a partir de las diferencias de las medias de los grupos y la desviación estándar ponderada, $d=(M1-M2)/DE$. En este sentido, valores de la d de Cohen inferiores a 0,20, señalan la no existencia de efecto; valores entre 0,21 a 0,49 hacen referencia a un pequeño efecto; así mismo, valores oscilantes entre 0,50 a 0,70 indican un moderado efecto; finalmente, valores mayores a 0,80 señalan un efecto grande⁵.

Con el objetivo de comparar la calidad de vida asociada con la salud según el sexo y tipo de enfermedad, López Cortés y col¹, usaron la U de Mann-Whitney, pero al no estar disponibles los datos de mediana y rango intercuartílico pero si la media, desviación estándar y el tamaño de cada grupo en la tabla 2 (p. 127), se calculó la d de Cohen, como una medida del tamaño del efecto para la comparación de dos grupos independientes. En el caso de la comparación por sexo, en ninguna de las dimensiones las diferencias son consideradas con un tamaño del efecto grande ($d_{Cohen} = > 0,80$), pero si con tamaños del efecto moderado, como en el caso de síntomas digestivos ($d_{Cohen} = 0,67$), sistémicos ($d_{Cohen} = 0,78$), función emocional ($d_{Cohen} = 0,65$) y función social ($d_{Cohen} = 0,71$). En relación a la comparación por tipo de enfer-

An. Sist. Sanit. Navar. 2016; 39 (3): 459-461

1. Universidad Privada del Norte. Lima. Perú.

2. Universidad del Pacífico. Lima. Perú.

Correspondencia:

Tomás Caycho

E-mail: tomas.caycho@upn.pe

medad, la situación es similar, ubicándose la mayoría en la categoría de moderado efecto (síntomas sistémicos, $d_{\text{Cohen}} = 0,72$; función emocional, $d_{\text{Cohen}} = 0,64$ y función social, $d_{\text{Cohen}} = 0,69$); mientras que la dimensión de síntomas digestivos ($d_{\text{Cohen}} = 0,47$) presenta un pequeño efecto. Los hallazgos del tamaño del efecto revelan que una gran cantidad de comparaciones se ubican en la categoría de magnitud pequeña y moderada, lo que puede generar errores en la interpretación de los resultados en base al valor de p .

Los resultados anteriores pueden explicarse por el procedimiento para el cálculo del tamaño del efecto empleado y la ausencia de un análisis de invarianza factorial. En el primer caso, si bien López Cortés y col¹ presentan el uso de la media y la desviación estándar, estos estadísticos se encuentran restringidos a distribuciones que satisfacen criterios de normalidad en el análisis de su distribución. A esta consideración, y a la mención de los autores acerca del empleo de estadísticos no paramétricos como la U de Mann-Whitney y el H de Kruskal-Wallis, se sugiere pertinente el uso de medidas robustas a la no normalidad, tales como la mediana y el rango intercuartílico, así como el empleo de la probabilidad de superioridad (PS) como índice del tamaño del efecto para comparaciones de dos grupos con medidas no paramétricas. La PS es definida como la probabilidad de que un puntaje elegido aleatoriamente de una población X sea superior a un puntaje elegida al azar de otra población Y^{6,7}. A continuación, se presenta su expresión matemática⁸:

$$PS_{\text{est}} = \frac{U}{mn}$$

donde: U es la Mann-Whitney, m es el número de participantes en el primer grupo; n es el número de participantes en el segundo grupo.

Esta medida de tamaño del efecto tiene la propiedad de ser de rápida interpretación, robusta a las violaciones de los supuestos paramétricos y no sensible a los valores extremos o *outliers*⁸. Dada la falta

de información, en la tabla 2 y en el mismo cuerpo del artículo, respecto a los valores de la U de Mann-Whitney, no es posible el cálculo del tamaño del efecto mediante este procedimiento. Aun así se sugiere su empleo en futuras investigaciones, que utilicen datos con distribuciones no paramétricas, como medidas del tamaño del efecto.

En el caso de la invarianza factorial o de medida, el estudio de López Cortés y col¹ carece de este análisis, requerido en todas aquellas investigaciones que tienen como objetivo comparar grupos. Mediante este procedimiento, se busca tener garantía de que los grupos a comparar entiendan de igual manera el constructo a analizar (en este caso, calidad de vida asociada a la salud); lo que permitiría concluir que la existencia (o no) de diferencias sería resultado de la existencia (o no) de diferencias reales respecto al constructo analizado⁹. De no ser así, las diferencias entre las medias presentarían errores potenciales¹⁰.

Los comentarios expuestos resaltan la importancia de presentar las medidas del tamaño del efecto acordes a los criterios de normalidad en los análisis de su distribución realizados, sugiriendo su implementación de manera cotidiana con el objetivo de brindar mayor evidencia, precisión y significado práctico a los resultados; además de ser medidas importantes para estudios de metaanálisis, destinados a la sistematización de resultados de estudios referidos a temas específicos en el ámbito de la salud.

BIBLIOGRAFÍA

1. LÓPEZ CORTÉS R, MARÍN FERNÁNDEZ B, HUESO MONTORO C, ESCALADA HERNÁNDEZ P, SANZ AZNAREZ A C, RODRÍGUEZ GUTIÉRREZ C. Calidad de vida relacionada con la salud en pacientes con enfermedad inflamatoria intestinal. *An Sist Sanit Navar* 2016; 39: 123-131.
2. DOMÍNGUEZ S. Diferencias entre grupos y magnitud del efecto: un análisis complementario. *Arch Argent Pediatr* 2016; 114: e300-e301.
3. FERGUSON CJ. An effect size primer: a guide for clinicians and researchers. *Prof Psychol Res Pr* 2009; 40: 532-538.

4. SULLIVAN G, FEINN R. Using effect size – or why the p value is not enough. *J Grad Med Educ* 2012; 4:279-282.
5. COHEN, J. *Statistical power analysis for the behavioral sciences*. (2ª ed.). Erlbaum, Hillsdale; 1998.
6. GRISSOM R, KIM J. *Effect sizes for research*. Lawrence Erlbaum Associates, editor. Mahwah, New Jersey; 2005.
7. RUSCIO J, MULLEN T. Confidence intervals for the probability of superiority effect size measure and the area under a receiver operating characteristic curve. *Multivariate Behavioral Res* 2012; 47: 201-223.
8. ERCEG-HURN D, MIROSEVICH V. Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *Am Psychol* 2008; 63: 591-601.
9. DOMÍNGUEZ-LARA S. Comparación del auto-concepto entre grupos, ¿sesgo o diferencias?: comentarios a Castillo et al. *Rev Chil Pediatr* 2016. Disponible en <http://www.sciencedirect.com/science/article/pii/S037041061630002X>
10. DRASGOW F. Biased test items and differential validity. *Psychol Bull* 1982; 92: 526-531.

