

Editorial

Medwave. Año XI, No. 8, Agosto 2011. Open Access, Creative Commons.

La interpretación de la magnitud de los efectos observados en los ensayos clínicos

Autor: Miguel Araujo Alonso⁽¹⁾

Filiación:

⁽¹⁾Hospital de Los Andes, Medwave

Correspondencia: maraujo@medwave.cl

doi: 10.5867/medwave.2011.08.5115

Ficha del Artículo

Citación: Araujo M. La interpretación de la magnitud de los efectos observados en los ensayos clínicos. *Medwave* 2011 Ago;11(08). doi: 10.5867/medwave.2011.08.5115

Fecha de envío: 27/7/2011

Fecha de aceptación: 27/7/2011

Fecha de publicación: 1/8/2011

Editorial

La falta de diferencia clínicamente significativa ha sido definida como aquel rango de valores que los clínicos, al comparar un nuevo tratamiento con la terapia estándar, consideran demasiado pequeño como para cambiar sus prácticas. Al contrario, una diferencia clínicamente relevante es aquella que lleva a adoptar la nueva práctica¹. Cualquiera sea el enfoque, en el contexto clínico la decisión no es simple porque además de la magnitud de la diferencia, entran en juego otros factores tales como el costo, la facilidad de administración, la necesidad de adquirir destrezas para administrar el nuevo tratamiento, los efectos adversos de una y otra intervención, la gravedad del problema clínico tratado, su vulnerabilidad a otras formas de terapia, entre otros. Así, determinar el rango de significación clínica es un juicio subjetivo y es difícil establecer un estándar universal al respecto. Su definición puede variar entre un médico y otro, así como entre la óptica de los profesionales y los pacientes.

Por ejemplo, van Walraven² evaluó mediante una encuesta la percepción de distintos especialistas (médico de familia, internista, cirujano vascular) sobre la diferencia mínima clínicamente importante, para el uso de propanolol con el objetivo de enlentecer el crecimiento de aneurismas aórticos abdominales, asumiendo que el fármaco fuera eficaz. El autor usó distintas medidas de resultado (velocidad de crecimiento o proporción de pacientes que requiere cirugía) y diferentes métodos de presentación de los datos (reducción absoluta de riesgo, número necesario a tratar). Las percepciones variaron considerablemente según especialidad, experiencia con la patología y con el uso de propanolol, y según el método usado para presentar los resultados. Con ese rango de variación, los tamaños de muestra (TM) que se requerirían para diseñar un estudio de efectividad oscilarían entre 100 y 3.000, es decir, el margen de "error" de estimación

(entendido como sobreestimar o subestimar el TM apropiado) es tremendamente amplio.

En el contexto de un ensayo clínico que comparó dos técnicas quirúrgicas en cáncer gástrico, una más extensa que la otra (D1 convencional vs. D2 radical), Fayers³ analizó las percepciones de los investigadores y cirujanos respecto a este punto. El tamaño de muestra del estudio se basó en el supuesto de una diferencia absoluta de un 13,5% en la supervivencia (20% para D1 y 33,5% para D2). En una encuesta a 26 cirujanos, la mayoría consideró que las terapias no serían significativamente distintas si la diferencia se encontrara entre 5% y 10%, algunos entre 10% y 15%, y para unos pocos debía ser de 20% o más. Para aquellos involucrados directamente en el estudio, el rango fue menor (2%-7%). Así, no sólo son esperables distintas opiniones entre los clínicos, sino también entre éstos y los investigadores que en definitiva deben justificar la importancia o necesidad del estudio.

Para algunas mediciones genéricas como calidad de vida, se aprecian también distintos criterios según el tipo de enfermedad que se esté analizando, o de acuerdo a cuál es el dominio específico de la escala que se esté analizando⁴.

También hay autores que han sugerido involucrar a los pacientes o representantes del público en la definición del efecto clínicamente importante de un nuevo tratamiento⁵, aunque la experiencia al respecto es poca. Frost publicó una investigación sobre el tema referida a las mediciones de calidad de vida⁶, y Barrett⁷ intentó establecer umbrales de significación clínica mínima para 4 intervenciones destinadas a tratar el resfriado común, mediante entrevistas cara a cara o telefónicas a 253 pacientes con síntomas respiratorios agudos, y encontró que éstos esperaban una reducción mínima promedio en la severidad de los síntomas que varió entre 25% y 57% según el tipo de

tratamiento, para justificar el costo y potenciales riesgos de los tratamientos. Observó también que la variabilidad en las respuestas entre individuos era muy amplia.

Chan⁸ propone atributos que considera relevantes para interpretar la importancia clínica de los resultados: identificación clara del desenlace primario, reporte de la diferencia que se espera encontrar entre los grupos, la significación estadística de los resultados, la presentación de los intervalos de confianza, y que exista una interpretación del autor respecto a la relevancia de los hallazgos.

En variables continuas, algunos utilizan como referencia los criterios de Cohen, quien propone interpretar como efecto "mediano" aquel detectable "a ojo desnudo" o perceptible en la práctica diaria sin necesidad de medirlo formalmente, "grande" aquel de magnitud tal que probablemente no se justifique hacer una investigación para demostrar algo tan evidente, y "pequeño" aquel que, no cumpliendo las condiciones de uno mediano, sí es lo suficientemente importante para justificar el estudio. Los tamaños de efecto (medida utilizada en algunos meta-análisis y ensayos originales) pequeños, medianos y grandes propuestos por Cohen son 0,2, 0,5 y 0,8 respectivamente⁹.

En resumen, la definición de significación clínica puede variar considerablemente entre individuos, investigadores, la población en general y profesionales asistenciales. Como sea, cuando usted revise un artículo de Medwave y se enfrente a interpretar la magnitud de un resultado, recuerde al menos tres cosas:

- Que un resultado sea estadísticamente significativo no implica que sea clínicamente importante
- Quien mejor puede juzgar lo que es importante o no quizás sea el paciente, entonces concéntrese en los resultados que él es capaz de percibir
- Observe con respeto la opinión del autor del estudio, pero saque sus propias conclusiones

Referencias

1. Simon SD. Interpreting negative studies. *J Androl.* 2001 Jan-Feb;22(1):13-6 ↑ | [PubMed](#) |
2. van Walraven C, Mahon JL, Moher D, Bohm C, Laupacis A. Surveying physicians to determine the minimal important difference: implications for sample-size calculation. *J Clin Epidemiol.* 1999 Aug;52(8):717-23. ↑ | [CrossRef](#) |
3. Fayers PM, Cuschieri A, Fielding J, Craven J, Uscinka B, Freedman LS. Sample size calculation for clinical trials: the impact of clinician beliefs. *Br J Cancer.* 2000;82:213-219. ↑ | [CrossRef](#) | [PubMed](#) | [PubMed Central](#) |
4. Wyrwich KW, Tierney WM, Babu AN, Kroenke K, Wolinsky FD. A comparison of clinically important differences in health-related quality of life for patients with chronic lung disease, asthma, or heart disease. *Health Serv Res.* 2005 Apr;40(2):577-91. ↑ | [CrossRef](#) | [PubMed](#) | [PubMed Central](#) |
5. Naylor CD, Llewellyn-Thomas HA. Can there be a more patient-centred approach to determining clinically important effect sizes for randomized treatment trials? *J Clin Epidemiol.* 1994 Jul;47(7):787-95. ↑ | [CrossRef](#) |
6. Frost MH, Bonomi AE, Ferrans CE, Wong GY, Hays RD; Clinical Significance Consensus Meeting Group. Patient, clinician, and population perspectives on determining the clinical significance of quality-of-life scores. *Mayo Clin Proc.* 2002 May;77(5):488-94. ↑ | [CrossRef](#) | [PubMed](#) |
7. Barrett B, Harahan B, Brown D, Zhang Z, Brown R. Sufficiently important difference for common cold: severity reduction. *Ann Fam Med.* 2007 May-Jun;5(3):216-23. ↑ | [CrossRef](#) | [PubMed](#) | [PubMed Central](#) |
8. Chan KB, Man-Son-Hing M, Molnar FJ, Laupacis A. How well is the clinical importance of study results reported? An assessment of randomized controlled trials. *CMAJ.* 2001 Oct 30;165(9):1197-202. ↑ | [PubMed](#) | [PubMed Central](#) |
9. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* 2007 May;39(2):175-91. ↑ | [CrossRef](#) | [PubMed](#) |



Esta obra de Medwave está bajo una licencia Creative Commons Atribución-NoComercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, Medwave.

