

Significancia estadística e importancia clínica de los resultados en los ensayos clínicos: un estudio metaepidemiológico

Juanita Sánchez Cano¹
Jorge Alejandro Pamplona Tobón²
Laura Jaramillo Santacoloma³
Jorge Hernando Donado Gómez⁴

Resumen

Objetivo: Determinar la significancia epidemiológica y clínica de los resultados para el desenlace primario de ensayos clínicos controlados aleatorizados, publicados en revistas médicas de alto impacto. **Metodología:** Se llevó a cabo un estudio observacional metaepidemiológico, en el que se aplicó un muestreo probabilístico para seleccionar 100 ensayos clínicos controlados aleatorizados, publicados en un periodo de 3 años (desde enero 2020 a diciembre 2022) en cuatro revistas médicas, y se clasificó el desenlace primario de cada ensayo clínico en cinco categorías según la significancia estadística y la relevancia clínica. **Resultados:** Se revisó un total de 100 artículos, se clasificaron en grupos, encontrándose que, de ellos, los desenlaces primarios de mayor frecuencia fueron estadísticamente no significativo y clínicamente no relevante (36 %), y estadísticamente significativo con relevancia

clínica no clara o incierta (28 %). El desenlace de menor frecuencia fue estadísticamente significativo y clínicamente no relevante (2 %). **Conclusión:** Este estudio destaca la importancia de una adecuada interpretación de los estudios científicos, donde se diferencien los términos de significancia estadística y relevancia clínica. Es importante evaluar la calidad de los estudios que soportan decisiones en la práctica clínica, ya que, según los resultados, la mayoría de los estudios analizados no tenían ni valor estadístico ni clínico. Hay pocos estudios metaepidemiológicos relacionados con este tema, por lo que se requieren mayores esfuerzos de investigación que permitan la comparación de resultados.

Palabras clave: relevancia clínica, análisis estadístico, investigación epidemiológica, ensayo clínico controlado aleatorizado, epidemiología analítica. (DeCS).

¹ Médico general. Hospital Pablo Tobón Uribe, Departamento de Hospitalización, Medellín, Colombia. juanita.sanchezc@upb.edu.co  ORCID: 0009-0003-0133-4088. **Google Scholar**

² Médico general. Hospital Pablo Tobón Uribe, Departamento de Hospitalización, Medellín, Colombia. jorge.pamplona@upb.edu.co  ORCID: 0000-0003-1630-2292. **Google Scholar**

³ Médico general. Hospital Pablo Tobón Uribe, Departamento de Hospitalización, Medellín, Colombia. laura.jaramillosantacoloma@upb.edu.co  ORCID: 0000-0002-3102-4510. **Google Scholar**

⁴ Profesor Titular Universidad Pontificia Bolivariana. Médico internista. Magíster en Epidemiología Clínica. Hospital Pablo Tobón Uribe, Medellín, Colombia. jdonado@hptu.org.co  ORCID: 0000-0001-8581-2267. **Google Scholar**

Statistical Significance and Clinical Importance of Results in Clinical Trials: a Meta-epidemiological Study

Abstract

Objects: To determine the epidemiological and clinical significance of results for the primary outcome of randomized controlled clinical trials published in high impact medical journals.

Methods: A meta-epidemiological observational study was conducted, in which probability sampling was applied to select 100 randomized controlled clinical trials, published over a 3-year period (from January 2020 to December 2022) in four medical journals (NEJM, JAMA, The Lancet and BMJ) and the primary outcome of each clinical trial was classified into five categories according to statistical significance and clinical relevance, using the 95% confidence interval.

Results: A total of 100 articles were reviewed and classified in groups, of which the most frequent primary outcomes were statistically non-significant and clinically irrelevant (36%), and statistically significant with unclear or uncertain clinical relevance (28%). The least frequent outcome was statistically significant and clinically not relevant (2%).

Conclusion: This study highlights the importance of an adequate interpretation of scientific studies, where the terms of statistical significance and clinical relevance are differentiated. It is important to evaluate the quality of the studies that support decisions in clinical practice, since according to the results most of the studies analyzed had neither statistical nor clinical value. There are few meta-epidemiological studies related to this topic, so greater research efforts are required to allow comparison of results.

Key words: clinical relevance, statistical analysis, epidemiological research, randomized controlled clinical trial.

Introducción

Los ensayos clínicos aleatorios (ECA) son ampliamente reconocidos como el mejor diseño epidemiológico para demostrar los beneficios y la seguridad de intervenciones en salud, incluyendo procedimientos quirúrgicos (Itani y Kibbe, 2022). Los resultados se pueden expresar mediante valores de p para probar pruebas de hipótesis o estimación puntual y por intervalo de confianza de las medidas de asociación. Las pruebas de significancia estadística han ocupado un lugar central en la investigación científica a lo largo de la historia. Durante años, estas técnicas fueron consideradas como el estándar, capaces de determinar, en conjunto con otros componentes como la metodología y el control de sesgos, la validez de un estudio.

Sin embargo, en las últimas décadas ha surgido un intenso debate en torno a su utilidad, especialmente cuestionando el valor de P (Clark, 2004), que representa la probabilidad de observar una diferencia en el estadístico igual o mayor al observado, bajo el supuesto de que no hay efecto (hipótesis nula). Se ha planteado que este valor ofrece una información limitada y, en ocasiones, poco relevante para los investigadores (Wang y Long, 2022; Greenland *et al.*, 2016).

En contraposición, se ha reconocido cada vez más la importancia de complementar o incluso reemplazar las pruebas de significancia con otras pruebas analíticas, como el intervalo de confianza. Este cambio de enfoque se ha gestado desde mediados del siglo pasado,

buscando proporcionar una comprensión más profunda y completa de los resultados de investigación (Gil y Castañeda, 2005; Naimi y Whitcomb, 2020).

Entre los argumentos esgrimidos por diversos autores, para evitar el uso exclusivo de las pruebas de hipótesis, se destaca la falta de información sobre la magnitud y precisión de los efectos observados, así como la limitada capacidad para comprender a fondo el fenómeno estudiado y la confiabilidad de los resultados obtenidos. Además, la terminología utilizada en el contexto de la significancia estadística puede llevar a interpretaciones erróneas, ya que no necesariamente implica relevancia clínica (Ciapponi, 2014; Hespagnol *et al.*, 2019).

En contraste, los procedimientos analíticos, como el intervalo de confianza, buscan no solo identificar la significancia práctica de los resultados, sino también impulsar la replicación y el desarrollo de nuevos estudios, contribuyendo así al avance del conocimiento científico y clínico (Schober y Vetter, 2020).

El presente estudio tiene como objetivo determinar la significancia epidemiológica y clínica de ensayos clínicos controlados aleatorizados, publicados en revistas médicas de alto impacto. Este enfoque busca ofrecer una visión integral de la investigación médica, promoviendo una evaluación más completa y rigurosa de la evidencia científica disponible.

Metodología

Se llevó a cabo un estudio observacional metaepidemiológico, en el que se aplicó un muestreo probabilístico para seleccionar 100 ensayos clínicos controlados aleatorizados (ECA). El tamaño de la muestra se estimó de acuerdo con el estudio de Johnston *et al.*, (2019) para estudios descriptivos y con una prevalencia esperada de 0.10 y una precisión

del intervalo de confianza 0.06 (Johnston *et al.*, 2019). Los criterios de inclusión abarcaron ECA de grupos paralelos, con diseño de superioridad y desenlace primario dicotómico, analizados mediante medidas de asociación como hazard ratio (HR) o riesgo relativo (RR), y en algunos casos odds ratio (OR), publicados en un periodo de 3 años (2020-2022). Se seleccionaron ECA publicados en cuatro revistas médicas generales de alto factor de impacto: New England Journal of Medicine (NEJM), The Journal of the American Medical Association (JAMA), The Lancet y British Medical Journal (BMJ). Los criterios de exclusión contemplaron ECA de no-inferioridad, asignación por grupos (conglomerados) o con grupos cruzados.

El análisis estadístico se realizó mediante frecuencias absolutas y relativas para variables cualitativas, y medidas de tendencia central y dispersión para variables cuantitativas. Se evaluaron variables como incidencia esperada y encontrada en grupo control e intervención, así como la diferencia absoluta y relativa esperada y encontrada en los resultados de los estudios. Se clasificó el desenlace primario de cada ECA en cinco categorías, según la significancia estadística y la relevancia clínica, utilizando el intervalo de confianza al 95 % (IC95 %) así (Itani y Kibbe, 2022): la diferencia mínima clínica esperada se tomaba del cálculo del tamaño de muestra, según lo que esperaban los autores que fuese el efecto de la intervención evaluada en el ECA.1

1. “Estadísticamente significativo y clínicamente relevante”: El IC95 % no cruzaba la unidad y el límite inferior, superaba la diferencia clínica mínima esperada.
2. “Estadísticamente significativo con relevancia clínica no clara o incierta”: El IC95 % no cruzaba la unidad, el límite inferior era menor a la diferencia clínica mínima esperada.
3. “Estadísticamente significativo y clínicamente no relevante”: El IC95 % no cruzaba la unidad, pero tanto el límite inferior como superior era menor a la diferencia mínima clínica

esperada.

4. *“Estadísticamente no significativo y clínicamente relevante”*: El IC95 % cruzaba la unidad, pero el límite inferior superaba (incluía o era compatible con) la diferencia mínima clínica esperada.
5. *“Estadística y clínicamente no relevante”*: El IC95 % cruzaba la unidad y el límite inferior era menor a la diferencia mínima clínica esperada.

Se emplearon el paquete estadístico Epidat versión 4.2. El presente estudio siguió los criterios establecidos para estudios metaepidemiológicos, adaptados de las *“Guidelines for Reporting Meta-epidemiological Methodology Research”* (Murad y Wang, 2017).

Debido a que el planteamiento de la investigación correspondió a un estudio metaepidemiológico, sin intervención en seres humanos, no fue necesario obtener aprobación de un comité de ética (World Medical Association, 2013).

Resultados

Entre los años 2020 - 2022 la búsqueda en PubMed encontró un total de 605 ensayos clínicos controlados aleatorizados (ECA), distribuidos en 176 de NEJM, 226 de JAMA, 118 de The Lancet y 85 de BMJ. De este total,

se seleccionaron mediante muestreo aleatorio simple 100 artículos que cumplieran con los criterios de elegibilidad establecidos (Figura 1). Se incluyeron 26 ensayos clínicos de NEJM, JAMA y The Lancet, y 22 de BMJ. Todos los artículos revisados fueron publicados entre 2020 y 2022, representando el 27 %, 28 % y 45 %; respectivamente para cada año.

En cuanto al tamaño muestral, se observó un promedio de 2991,3 participantes (SD 6130,8), rango (73 - 40872) participantes. La mediana fue de 754 participantes (RIC 391,5 - 2476,3) (Tabla 1).

Respecto a las incidencias encontradas en los grupos de intervención, se calculó una media de 25,46 (SD 20,1), rango (1,38 - 92,4). La mediana fue de 19,8 (RIC 15 - 40), y la diferencia relativa esperada media fue de 33,2 (SD 16,1), rango (1,2 - 75). La mediana de la diferencia relativa esperada fue de 30,1 (RIC 20 - 40).

En el análisis de la significancia estadística y clínica para el desenlace primario, se encontró que el resultado más común fue la falta de significancia estadística y relevancia clínica, con un 36 % de los casos (Tabla 2).

Se encontró que el promedio de la reducción relativa del riesgo fue de un 15 % con IC no significativo entre aumento del 19 % o una reducción del 39 % (Tabla 3).

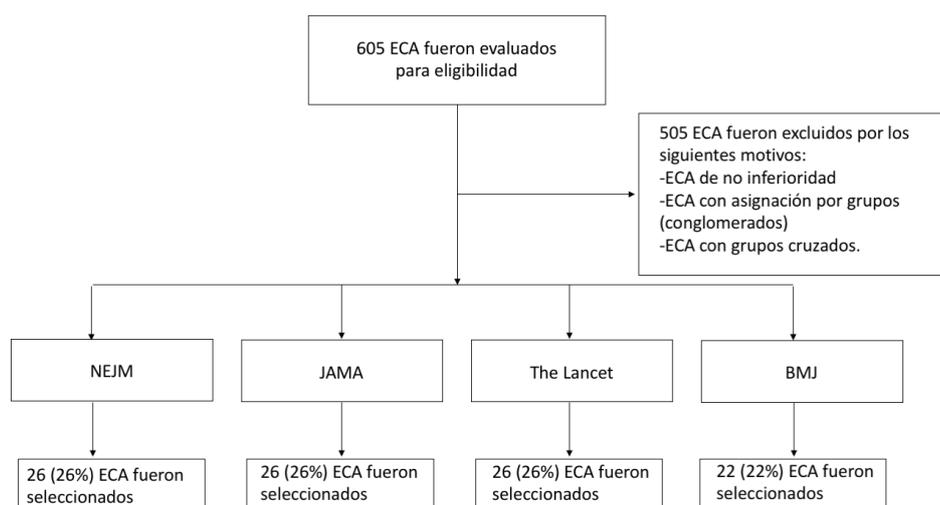


Figura 1. Diagrama de flujo del estudio.

Tabla 1. Características de los ECA analizados

Característica	n = 100 ECA
Tamaño de muestra	
Media (SD)	2991,3 (6130,8)
Mediana (P ₂₅ - P ₇₅)	754,0 (397,5 - 2476,3)
Mediana (P ₁₀ - P ₉₀)	754,0 (259,2 - 6407,3)
Mediana (min - máx.)	754,0 (73 - 40872)
Incidencia esperada (%) del desenlace en el grupo control	
Media (SD)	27,08 (18,3)
Mediana (P ₂₅ - P ₇₅)	21,6 (15 - 4)
Mediana (P ₁₀ - P ₉₀)	21,6 (6,8 - 50)
Mediana (min - máx.)	21,6 (0,6 - 85)
Incidencia esperada (%) del desenlace en el grupo de intervención	
Media (SD)	25,5 (20,8)
Mediana (P ₂₅ - P ₇₅)	19,80 (9,7 - 35)
Mediana (P ₁₀ - P ₉₀)	19,80 (4,1 - 59,8)
Mediana (min - máx.)	19,80 (1,4 - 89)
Incidencia encontrada (%) del desenlace en el grupo control	
Media (SD)	25,5 (18,9)
Mediana (P ₂₅ - P ₇₅)	20,6 (10,3 - 36,8)

Mediana ($P_{10} - P_{90}$)	20,6 (4,2 - 50,9)
Mediana (min - máx.)	20,6 (1,0 - 92,4)
Incidencia encontrada (%) del desenlace en el grupo de intervención	
Media (SD)	25,50 (20,1)
Mediana ($P_{25} - P_{75}$)	19,75 (11 - 41)
Mediana ($P_{10} - P_{90}$)	19,75 (3 - 53,8)
Mediana (min - máx.)	19,75 (0,2- 93,5)
Diferencia esperada absoluta (puntos porcentuales) del desenlace	
media (SD)	9,60 (7,5)
mediana ($P_{25} - P_{75}$)	8,75 (4 - 14,9)
mediana ($P_{10} - P_{90}$)	8,75 (1,6 - 16,7)
mediana (min - máx.)	8,75 (0,5 - 45)
Diferencia encontrada absoluta (puntos porcentuales) del desenlace	
media (SD)	6,1 (7,9)
mediana ($P_{25} - P_{75}$)	3,3 (1,1 - 8)
mediana ($P_{10} - P_{90}$)	3,3 (0,2 - 15,5)
mediana (min - máx.)	3,3 (0 - 47)
Diferencia relativa (%) esperada del desenlace	
media (SD)	33,2 (16,05)
mediana ($P_{25} - P_{75}$)	30,1 (20,0 - 40)
mediana ($P_{10} - P_{90}$)	30,1 (16,7 - 56,6)
mediana (min - máx.)	30,1 (1,2 - 75)
Diferencia relativa (%) encontrada del desenlace	
media (SD)	24,5 (23,2)
mediana ($P_{25} - P_{75}$)	15,1 (5,8 - 33,6)
mediana ($P_{10} - P_{90}$)	15,1 (1,6 - 62,9)
mediana (min - máx.)	15,1 (0 - 89)

Tabla 2. Tipos de desenlaces encontrados

Resultado	n (%)
Estadísticamente significativo y clínicamente relevante	9 (9)
Estadísticamente significativo con relevancia clínica no clara o incierta	28 (28)
Estadísticamente significativo y clínicamente no relevante	2 (2)
Estadísticamente no significativo y clínicamente relevante	25 (25)
Estadísticamente no significativo y clínicamente no relevante	36 (36)

Tabla 3. Resultados de 100 ensayos según el tipo de desenlaces, positivo o negativo

Tipo de desenlace (n = 100)		
Estimador puntual (RR, OR, HR)	Positivo (n = 28)	Negativo (n = 72)
media (SD)	2,3 (2,8)	0,9 (0,4)
mediana (min - máx.)	1,1 (0,5 - 11,10)	0,9 (0,1 - 3,4)
Límite inferior del IC	Positivo	Negativo
media (SD)	1,4 (1,3)	0,6 (0,3)
mediana (min - máx.)	0,9 (0,4 - 6,0)	0,6 (0,0 - 1,0)
Límite superior del IC	Positivo	Negativo
media (SD)	4,1 (6,2)	1,2 (0,7)
mediana (min - máx.)	1,6 (0,7 - 23,0)	1,1 (0,3 - 5,8)
Estimador puntual RRR	Positivo	Negativo
media (SD)	-130,2 (277,2)	14,9 (40,2)
mediana (min - máx.)	-10,00 (-1010,0 - 50,0)	10,00 (-240,0 - 90,0)
Límite inferior del RRR	Positivo	Negativo
media (SD)	-313,0 (617,8)	-18,6 (64,9)
mediana (min - máx.)	-60,0 (-2200,0 - 30,0)	-10,0 (-480,0 - 70,0)
Límite superior del RRR	Positivo	Negativo
media (SD)	-37,1 (132,2)	38,6 (28,6)
mediana (min - máx.)	10,0 (-500,0 - 60,0)	40,0 (0,0 - 100,0)

Discusión

El presente estudio abordó la relevancia y utilidad de los resultados de ensayos clínicos controlados aleatorizados de grupos paralelos de superioridad, centrándose en la significancia epidemiológica y clínica de los resultados obtenidos en un conjunto selecto de publicaciones médicas de alto factor de impacto.

Nuestros hallazgos reflejan una amplia distribución de ECA entre las principales revistas médicas, la selección aleatoria de 100 ECA permitió una representación diversa y robusta de las investigaciones clínicas recientes.

En cuanto al tamaño muestral y las incidencias encontradas en los grupos de intervención, se observó una variabilidad considerable, lo que sugiere una amplia gama de diseños de estudio y características de las poblaciones estudiadas en la literatura médica.

El análisis de la significancia estadística y clínica para el desenlace primario reveló una proporción significativa de resultados que no alcanzaron tanto la significancia estadística como la relevancia clínica. Este hallazgo respalda los argumentos previamente planteados en la introducción, sobre la limitación de las pruebas de significancia estadística, para proporcionar una evaluación completa y precisa del impacto clínico de una intervención (Ciapponi, 2014; Hespanhol *et al.*, 2019).

Este estudio respalda la creciente tendencia en la literatura científica hacia un enfoque más integral y crítico en la evaluación de la evidencia, que va más allá de únicamente la determinación de la significancia estadística. El énfasis en la significancia clínica y la interpretación contextual de los resultados puede proporcionar una comprensión más completa y pragmática de la eficacia y relevancia de las intervenciones médicas (Ciapponi, 2014).

En la literatura no se encontraron estudios previos o metaanálisis en los cuales se evalúe la significancia clínica o estadística, o ambas de un grupo de ECA, realizados bajo la metodología analítica propuesta en este estudio; por lo que los desenlaces encontrados no pueden ser comparados con otros hallazgos previos, evidenciando que se requiere de mayor investigación en el campo.

Lo anterior puede atribuirse a que es poco frecuente que los estudios publicados incluyan consideraciones sobre la magnitud del efecto que se considera clínicamente importante, y no existe consenso claro sobre los criterios a seguir para su establecimiento (Ochoa Sangrador, 2010). En la literatura se encontró que diversos autores han intentado definir el concepto de diferencia clínicamente importante, entendido para la mayoría como “significación clínica” y para el presente estudio como “relevancia clínica”. Por ejemplo, Hollon y Flick consideran que la unidad mínima de significación clínica debería ser definida en términos del más pequeño de los cambios válidos, que sea de interés para algunas pero no necesariamente todas las partes implicadas en un escenario clínico (Hollon y Flick, 1988). Así mismo, Lindgren *et al.*, (1993) señalan que, al comparar dos métodos de tratamiento, es clínicamente significativa la diferencia más pequeña respecto a una variable de resultado importante, que fundamentará una decisión de modificar un tratamiento. Y, por otro lado, algunos autores como Kingman (1992) establecen que la significación estadística debería ser considerada una condición necesaria para la “significación clínica” y que ambas significaciones, estadística y clínica, deberían coincidir. Sin embargo, se considera que para cumplir este requerimiento se vuelve necesario establecer una definición clara de significación clínica para poder ser extrapolada a los resultados de los diferentes estudios encontrados en la literatura.

La limitación más importante para el desarrollo de este estudio fue la ausencia de estudios

previos sobre el tema, que no permitió realizar comparaciones en los desenlaces. Finalmente, en el estudio solo se incluyeron artículos publicados en las cuatro revistas médicas consideradas de mayor relevancia, por lo que se desconoce cómo es el comportamiento de los tipos de desenlaces en los ECA publicados en otras revistas.

Por otro lado, como ventajas, se utilizaron criterios de inclusión claros para mejorar los resultados y su interpretación. Dado que no encontramos trabajos, investigaciones o iniciativas similares en la literatura, podría considerarse el primer trabajo investigativo sobre el tema y podrá servir como base para el desarrollo de futuras investigaciones.

Conclusiones

Determinar la magnitud de la diferencia que se considera clínicamente relevante es crucial para evaluar los resultados de un ECA. Por lo tanto, es importante reconocer que no se puede limitar los desenlaces a una interpretación puramente estadística de estos datos, ya que eso significaría pasar por alto los diversos elementos presentes en el contexto clínico que se tienen que abordar.

Nuestros hallazgos respaldan la necesidad de adoptar enfoques metodológicos más sofisticados y contextualizados en la evaluación de la evidencia científica, reconociendo tanto la importancia de las pruebas de significancia estadística como sus limitaciones inherentes. En este estudio se subraya la importancia de considerar múltiples dimensiones de la evidencia, incluida la relevancia clínica, en la toma de decisiones médicas y el avance del conocimiento científico.

Referencias

1. Ciapponi, A. (2014). ¿Significancia clínica o significancia estadística? *Evidencia Actualización en la Práctica Ambulatoria*, 16(4). <https://doi.org/10.51987/evidencia.v16i4.6208>
2. Clark, M. L. (2004). Los valores P y los intervalos de confianza: ¿en qué confiar? *Revista Panamericana de Salud Pública*, 15(5), 293-296.
3. Gil, J. F. y Castañeda, J. A. (2005). Una mirada al valor de p en investigación. *Revista Colombiana de Psiquiatría*, 34(3), 414-424.
4. Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N. y Altman, D. (2016). Statistical Tests, P values, Confidence Intervals, and Power: A Guide to Misinterpretations. *European Journal of Epidemiology*, 31, 337-350.
5. Hespanhol, L., Vallio, C. S., Costa, L. y Saragiotto, B. T. (2019). Understanding and Interpreting Confidence and Credible Intervals Around Effect Estimates. *Brazilian Journal of Physical Therapy*, 23(4), 290-301.
6. Hollon, S. D. y Flick, S. N. (1988). On the Meaning and Methods of Clinical Significance. *Behavioral Assessment*, 10(2), 197-206.
7. Itani, K. y Kibbe, M. (2022). Clinical Trials Design, Scientific Rigor, and Higher-Level Evidence in Surgery. *JAMA Surgery*, 157(12), 1078-1079.

8. Johnston, K. M., Lakzadeh, P., Donato, B. M. K. y Szabo, S. M. (2019). Methods of Sample Size Calculation in Descriptive Retrospective Burden of Illness Studies. *BMC Medical Research Methodology*, 19(1), 9. <https://doi.org/10.1186/s12874-018-0657-9>
9. Kingman, A. (1992). Statistical vs Clinical Significance in Product Testing: Can They Be Designed to Satisfy Equivalence? *Journal of Public Health Dentistry*, 52(6), 353-360.
10. Lindgren, B. R., Wielinski, C. L., Finkelstein, S. M. y Warwick, W. J. (1993). Contrasting Clinical and Statistical Significance within the Research Setting. *Pediatric Pulmonology*, 16(6), 336-340.
11. Murad, M. H. y Wang, Z. (2017). Guidelines for Reporting Meta-epidemiological Methodology Research. *Evidence Based Medicine*, 22(4), 139-142.
12. Naimi, A. I. y Whitcomb, B. W. (2020). Can Confidence Intervals Be Interpreted? *American Journal of Epidemiology*, 189(7), 631-633.
13. Ochoa Sangrador, C. (2010). Evaluación de la importancia de los resultados de estudios clínicos. Importancia clínica frente a significación estadística. *Evidencias en Pediatría*, 6(2), 15. <https://evidenciasenpediatria.es/articulo/5413/evaluacion-de-la-importancia-de-los-resultados-de-estudios-clinicos-importancia-clinica-frente-a-significacion-estadistica>
14. Schober, P. y Vetter, T. R. (2020). Confidence Intervals in Clinical Research. *Anesthesia and Analgesia*, 130(5),1303.
15. Wang, M. y Long, Q. (2022). Addressing Common Misuses and Pitfalls of P Values in Biomedical Research. *Cancer Research*, 82(15), 2674-2677.
16. World Medical Association. (2013). Ethical Principles for Medical Research Involving Human Subjects. *JAMA*, 310(20), 2191-2194.