

## Discrepancias entre las frecuencias de los desenlaces primarios propuestos y obtenidos en ensayos clínicos aleatorizados

### Differences between the proposed frequencies of the research protocols of randomized clinical trials with outcomes obtained at the end of the primary study outcomes

Mateo Zuluaga Gómez<sup>1</sup>, Daniel Vanegas Isaza<sup>1</sup>,  
Jorge Hernando Donado Gómez<sup>2</sup>

#### Resumen

**Objetivo:** Determinar las diferencias entre las frecuencias de los desenlaces propuestos en el protocolo inicial de los ensayos clínicos aleatorizados respecto a los desenlaces obtenidos al final del estudio.

**Metodología:** Estudio observacional, de corte transversal, que incluyó ensayos clínicos de cardiología publicados en The New England Journal of Medicine desde 2002 a 2013.

**Resultados:** Se identificaron 28 ensayos clínicos. En el 60,7 % de los estudios la incidencia estimada del desenlace fue mayor que la encontrada, y se presentó discrepancia (diferencia mayor al 10 %) en 13 (46.4 %) estudios. El 10,7 % de los estudios presentaba discrepancia entre el tamaño de muestra estimado por los autores y el calculado por nosotros.

**Conclusión:** Encontramos de que a pesar que los autores de los ensayos clínicos tienen un buen cálculo del tamaño de la muestra, sus supuestos no son buenos, ya que la incidencia esperada en el grupo control suele ser sobreestimada en el 60 % de los casos.

**Palabras clave:** ensayo clínico, tamaño de la muestra, desenlace, incidencia.

#### Abstract

**Objective:** To determine the differences between the frequencies of the outcomes proposed in the initial protocol of randomized clinical trials for the outcomes obtained at endpoint.

**Methods:** An observational, cross-sectional that included cardiology clinical trials published in The New England Journal of Medicine from 2002-2013.

**Results:** We identified 28 trials. In 60.7 % of the estimated outcome studies incidence was higher than that found, showing discrepancy (difference greater than 10 %) in 13 (46.4%)

<sup>1</sup> Estudiantes de medicina, Facultad de Medicina Universidad Pontificia Bolivariana.

<sup>2</sup> Médico internista y epidemiólogo clínico Hospital Pablo Tobón Uribe. Profesor titular Universidad Pontificia Bolivariana.

**Correspondencia:** Jorge Hernando Donado Gómez. Calle 78 B n° 69 - 240. Hospital Pablo Tobón Uribe, noveno piso norte. Medellín (Antioquia, Colombia). Teléfono fijo: 574 - 4459753 Celular: 57 - 313 685 3078. jdonado@hptu.org.co

*studies. 10.7 % of the studies showed a discrepancy between the size estimated by the authors and the sample calculated by us.*

**Conclusion:** *We found that although the authors of clinical trials have a good estimate of the sample size, its assumptions are not good, because the expected incidence in the control group is often overestimated in 60 % of cases.*

**Keywords:** clinical trial, sample size, outcome, incidence.

## INTRODUCCIÓN

Los ensayos clínicos son estudios que tienen como objetivo evaluar la eficacia y seguridad de diversas intervenciones en salud, para dar respuesta a problemas de interés (1).

En la planeación del ensayo clínico de superioridad se debe establecer un tamaño de muestra que permita identificar las diferencias que se esperan encontrar con la intervención(2).

Existen factores que son de orden estadístico, los cuales condicionan el cálculo del tamaño de muestra (3-8).

- Error  $\alpha$ , corresponde a la probabilidad de rechazar la hipótesis nula cuando en realidad es cierta. Es un falso positivo. Lo más frecuente es aceptar un riesgo  $\alpha \leq 0,05$ .
- Error  $\beta$ , es la probabilidad de error al no rechazar la hipótesis nula cuando en realidad es falsa. Es un falso negativo. Los valores que más se aceptan son un  $\beta$  entre 0,10 y 0,20.
- El poder estadístico ( $1 - \beta$ ) es la probabilidad de rechazar la hipótesis nula cuando es falsa.
- Incidencia en grupo no expuesto. Es la frecuencia esperada del evento que se va a estudiar en el grupo control.
- La diferencia del efecto que se busca encontrar entre los grupos estudiados en el desenlace de interés.

El cálculo de los tamaños de muestras y los análisis estadísticos son de gran importancia en la planeación, interpretación y conclusiones de los ensayos clínicos con asignación aleatoria (9-12).

En los resultados de los ensayos clínicos se puede encontrar resultados estadísticamente no significativos, que pueden representar falta de poder, ya que los supuestos para calcular el tamaño de la muestra pueden ser sobreestimados con los encontrados (8,9,11).

El objetivo de esta investigación fue comparar los supuestos para calcular el tamaño de la muestra con los valores reales encontrados en el estudio, determinar el grado de discrepancia que se encuentra y cómo afecta el desenlace primario propuesto.

## METODOLOGÍA

Se realizó un estudio observacional, de corte transversal, metaepidemiológico.

Como criterios de selección se incluyeron ensayos clínicos controlados con asignación aleatoria de grupos paralelos en el área cardiovascular (incluyendo patologías endocrinológicas como diabetes mellitus) publicados en *The New England Journal of Medicine* (NEJM) antes de mayo de 2013, que tuvieran desenlace primario dicotómico y que fueran estudios de superioridad. Se realizó un muestreo no probabilístico, por conveniencia.

De cada uno de los artículos se revisó el protocolo, análisis estadístico, desenlaces, incluyendo el error alfa, beta y poder; así mismo, la incidencia del evento esperada en el grupo control y la incidencia encontrada en dicho grupo en los resultados. Se incluyó también el tamaño de muestra estimado por los investigadores en el protocolo y el tamaño de muestra calculado a través de la fórmula (ver figura 1) (3).

$$N = (Z_{2\alpha} + Z_{\beta})^2 \times S^2 / \Delta^2$$

**Figura 1.** Fórmula de estimación del tamaño de muestra

N: Número necesario de pacientes por grupo

$Z_{2\alpha}$ : Constante para cada valor de  $2\alpha$

$Z_{\beta}$ : Constante para cada valor  $\beta$

S: Varianza

$\Delta$ : Diferencia esperada entre la intervención y el grupo control

Finalmente se tomó la diferencia relativa esperada por encontrar de acuerdo con el evento y la diferencia obtenida en los resultados. Posteriormente los datos fueron digitados en una base de datos en Microsoft Excel 2010. Para garantizar la calidad del dato, dos autores realizaron revisión aleatoria del 20 % de la información de cada uno de los autores. Se consideró una diferencia  $\geq 10$  % entre las incidencias del grupo control calculados en el protocolo y los encontrados en los resultados como definición de discrepancia (13).

Los resultados se expresaron de acuerdo con la variable como frecuencias absolutas y relativas para las cualitativas, y como media y desviación estándar o mediana y rango intercuartílico para las variables cuantitativas.

## RESULTADOS

Se identificaron en total 28 ensayos clínicos con asignación aleatoria publicados desde 2002 hasta 2013 en el *NEJM* que cumplieran con los criterios de selección (14-41).

Todos los artículos presentaban un error alfa de 0,05. Con respecto al error beta, el más frecuente encontrado fue de 0,20 (57,61%). Las demás características estadísticas del tamaño de muestra pueden observarse en la tabla 1.

La mediana de la incidencia del evento esperada fue 0.25 RIQ (0.17-0.40), rango [0.01-0.85], y la encontrada fue 0.21 RIQ (0.11-0.38), rango [0.01 -0.7]; diferencia 0.02 IC 95% (-0.08 a 0.12), valor p 0.47. La mediana del tamaño de la muestra calculado en los protocolos fue 1.636 RIQ (947 - 4.050), rango [74-12.500], y la calculada 1.580 RIQ (795-3.677), rango [80-11.236]; diferencia 249 IC 95% (-1209 a 1707), valor p 0.62 (ver tabla 2).

**Tabla 1.** Características estadísticas del cálculo del tamaño de muestra en 28 ensayos clínicos

|                   | N (%)    | IC 95%         |
|-------------------|----------|----------------|
| <b>Error alfa</b> |          |                |
| 0.05              | 28 (100) |                |
| <b>Poder</b>      |          |                |
| 80%               | 16(57.1) | (38,81 – 75,4) |
| 82%               | 1(3.5)   | (0 – 10,4)     |
| 84%               | 1(3.5)   | (0 – 10,4)     |
| 90%               | 10(35.7) | (17,96 – 53,4) |
| <b>2 colas</b>    | 28 (100) |                |

**Fuente:** Datos tabulados por autores.

**Tabla 2.** Resultados de pruebas estadísticas entre tamaños de muestras y diferencias relativas calculados en los protocolos de los ensayos clínicos comparado con los calculados por los investigadores

|   | Mediana | Media | SD    | Q1    | Q3  | Max  | Min   | Número de Discrepancias N(%) |
|---|---------|-------|-------|-------|-----|------|-------|------------------------------|
| Diferencia relativa entre incidencia esperada y encontrada                | 0,02    | 0,03  | 0,1   | 0,00  | 0,1 | 0,2  | -0,2  | 13 (46.4%)                   |
| Diferencia relativa de las diferencias entre grupos esperada y encontrada | 0,1     | 0,05  | 0,2   | -0,01 | 0,1 | 0,3  | -0,5  | 19 (67.9%)                   |
| Diferencia absoluta entre los tamaños de muestra estimado y calculado     | 133     | 248,8 | 584,3 | 11,5  | 376 | 1362 | -1142 | 3(10.7%)                     |

SD: Desviación estándar Q1: Cuartil 1 (25 %); Q3: Cuartil 3 (75 %); Max: Valor máximo Min Valor. mínimo.

**Fuente:** Datos tabulados por autores.

En 3 de 28 estudios (10,7%) se encontró discrepancia entre el tamaño de muestra estimado por los autores y el calculado por nosotros (15,19,25). En ninguno de los tres estudios se encontraron diferencias estadísticamente significativas en sus resultados.

En 17 de 28 estudios (60,7 %) la incidencia estimada fue mayor que la encontrada, y presentó discrepancia en 13 estudios (46.4 %) (17, 18, 20, 22, 24-27, 29, 30, 33, 35, 37). En 7 de estos 13 estudios (54 %) los resultados no fueron estadísticamente significativos.

En 19 de 28 estudios (67,9 %) hubo una diferencia menor al 90 % entre el valor encontrado y el esperado con respecto a la diferencia por detectar (15, 17, 19, 21, 22, 24, 25, 27-32, 34, 36, 38-41).

## DISCUSIÓN

Los resultados de los ensayos clínicos pueden verse afectados por el tipo de análisis escogido o por el tamaño de muestra calculado. Cuando alguna de estas dos características se altera se pueden hacer interpretaciones erróneas de los resultados (13).

Este estudio muestra que la mayoría de los autores hace un buen estimado del tamaño de muestra previo al inicio de sus ensayos clínicos. Sin embargo, en cerca del 11 % el tamaño de muestra del protocolo fue menor al 90% del calculado (discrepancia mayor al 10 %), lo cual podría explicar que en dichos ensayos no se alcanzaron los resultados deseados por baja potencia (11).

Respecto a las incidencias del evento en el grupo control de cada ensayo clínico, se encontró que en el 46,4 % había discrepancia, y de estos estudios, en más de la mitad no lograron resultados estadísticamente significativos.

El sobreestimar la incidencia del desenlace primario en el grupo control puede explicarse por la vigencia temporal de los artículos de referencia o el tipo de población seleccionado, los cuales pueden variar la incidencia del evento.

En 2/3 de los estudios la reducción esperada del evento con la intervención tuvo una discrepancia >10 % de la encontrada, lo cual se explica, al menos en parte, por la sobreestimación del desenlace en el grupo control, que ocurrió en el 60.7 %.

Entre las fortalezas de este estudio se plantea una evidencia empírica sobre las discrepancias en la estimación del tamaño de muestra por parte de los investigadores, el incluir un número importante de estudios.

La limitación de esta investigación consistió en que en nuestros criterios de selección solo incluimos artículos del área de cardiología y una revista con alto factor de impacto, lo cual no puede extrapolarse a investigadores que realizan estudios y publican en otras revistas de menor categoría.

Los investigadores deben vigilar de forma periódica la validez de sus datos a través de un comité de monitoreo y seguridad de datos el avance de los estudios, y en caso de ser necesario realizar ajustes, además de garantizar la validez científica e integridad de la información; principalmente en aquellos estudios que tengan largos períodos de seguimiento o que traten con población vulnerable o de riesgo (42-44).

## CONCLUSIONES

En resumen, encontramos que a pesar de que los autores de los ensayos clínicos tienen un buen cálculo del tamaño de la muestra, sus supuestos no son buenos, ya que la incidencia esperada en el grupo control suele ser sobreestimada en el 60 % de los casos y se encuentra discrepancia en un 46.4 %. Esto implica que los investigadores deben vigilar de forma periódica, a través de un comité de monitoreo y seguridad de datos, el avance de los estudios, y en caso necesario realizar ajustes al tamaño de la muestra, con base en los datos reales que están consiguiendo durante el estudio, con miras a mantener el poder adecuado para poder detectar las diferencias esperadas.

**Conflicto de intereses:** ninguno.

**Financiación:** Universidad Pontificia Bolivariana

## REFERENCIAS

1. Laponte JR. El ensayo clínico controlado. *Principios básicos de investigación clínica* 2001; 27-54.
2. Chan A, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA et al. *Research methods and reporting spirit 2013 explanation and elaboration: guidance for protocols of clinical trials* 2013; 1-42.
3. Christensen E. Methodology of superiority vs. equivalence trials and non-inferiority trials. *Journal of hepatology* 2007;46(5):947-54. doi:10.1016/j.jhep.2007.02.015
4. Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: basic principles and common pitfalls. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association* 2010;25(5):1388-93. doi:10.1093/ndt/gfp732.
5. Charles P, Giradeau B, Dechartres A, Baron G, Ravaud P. *Reporting of sample size calculation in randomised controlled trials: review*. *BMJ (Clinical research ed.)*; 2009. p. 338. doi:10.1136/bmj.b1732.
6. Marrugat J, Vila J, Pavesi M. *Estimación del tamaño de la muestra en la investigación clínica y epidemiológica*; 1998. p. 267-76.
7. Zhong B. How to calculate sample size in randomized controlled trial? *Journal of thoracic disease* 2009;1(1):51-4.
8. Florey C V. Sample size for beginners. *BMJ*. 1993;306(306):1181-4.
9. Altman, D G, Schulz KF, Moher D, Egger M, Davidoff F et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
10. Altman, D G, Moher DSK. Peer review of statistics in medical research: reporting power calculations is important. *BMJ* 2002;(325):491.

11. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA: the journal of the American Medical Association* 1994;272:122-4.
12. Weaver CS, Leonardi-Bee J, Bath-Hextall FJ BP. Sample size calculations in acute stroke trials: a systematic review of their reporting, characteristics, and relationship with outcome. *Stroke* 2004;35:1216-24. doi:10.1161/01.STR.0000125010.70652.93.
13. Chan AW, Hróbjartsson A, Jørgensen KJ, Gøtzsche PC et al. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ* 2008;4:337.
14. Imazio M, Brucato A, Cemin R, Ferrua S, Maggolini S, Beqaraj F et al. A randomized trial of colchicine for acute pericarditis. *The New England journal of medicine* 2013;369(16):1522-8. doi:10.1056/NEJMoa1208536.
15. Zeeuw D De, Eckardt K, Feyzi JM, Ivanovich P, McMurray JJV et al. A trial of Darbepoetin Alfa in Type 2 Diabetes and Chronic Kidney Disease. *The New England journal of medicine* 2009;2019-32. doi:10.1056/NEJMoa0907845.
16. Steg PG, Van 't Hof A, Hamm CW, Clemmensen P, Lapostolle F, Coste P et al. Bivalirudin started during emergency transport for primary PCI. *The New England journal of medicine* 2013;369(23):2207-17. doi:10.1056/NEJMoa1311096.
17. Wing RR, Bolin P, Brancati FL, Bray G a, Clark JM, Coday M et al. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. *The New England journal of medicine* 2013;369(2):145-54. doi:10.1056/NEJMoa1212914.
18. Marchioli R, Finazzi G, Specchia G, Cacciola R, Cavazzina R, Cilloni D et al. Cardiovascular events and intensity of treatment in polycythemia vera. *The New England journal of medicine* 2013;368(1):22-33. doi:10.1056/NEJMoa1208500.
19. Chochrad D, Aldecoa C, Brasseur A, Defrance P, Gottignies P, Vincent J et al. Comparison of dopamine and norepinephrine in the treatment of shock. *The New England journal of medicine* 2010;779-89. doi:10.1056/NEJMoa0907118.
20. Wang Y, Wang Y, Zhao X, Liu L, Wang D, Wang C et al. Clopidogrel with aspirin in acute minor stroke or transient ischemic attack. *The New England journal of medicine* 2013;369(1):11-9. doi:10.1056/NEJMoa1215340.
21. Wiedemann HP, Wheeler AP, Bernard GR, Thompson BT, Hayden D, deBoisblanc B et al. Comparison of two fluid-management strategies in acute lung injury. *The New England journal of medicine* 2006;354(24):2564-75. doi:10.1056/NEJMoa062200.
22. Douglas IS, Finfer S, Gårdlund B, Marshall JC, Rhodes A, Williams MD et al. Drotrecogin alfa (activated) in adults with septic shock. *The New England journal of medicine* 2012;2055-64. doi:10.1056/NEJMoa1202290.
23. Álvarez-Ortega C. Early surgery versus conventional treatment for infective endocarditis. *The New England journal of medicine* 2012;212(10):501. doi:10.1056/NEJMoa1112843.
24. Cook D, Meade M, Guyatt G, Walter S, Heels-Ansdell D, Warkentin TE et al. Dalteparin versus unfractionated heparin in critically ill patients. *The New England journal of medicine* 2011;364(14):1305-14. doi:10.1056/NEJMoa1014475.
25. Benavente OR, Hart RG, McClure L, Szychowski JM, Coffey CS, Pearce L a. Effects of clopidogrel added to aspirin in patients with recent lacunar stroke. *The New England journal of medicine* 2012;367(9):817-25. doi:10.1056/NEJMoa1204133.
26. De Bruyne B, Pijls NHJ, Kalesan B, Barbato E, Tonino P, Piroth Z et al. Fractional flow reserve-guided PCI versus medical therapy in stable coronary disease. *The New England journal of medicine* 2012;367(11):991-1001. doi:10.1056/NEJMoa1205361.
27. Amarenco P, Bogousslavsky J, Callahan A, Gold - LB, Hennerici M, Rudolph AE et al. High-dose atorvastatin after stroke or transient ischemic attack. *The New England*

- journal of medicine* 2006;549-59. doi:10.1056/NEJMoa061894.
28. Heart T, Prevention O, Hope E. Homocysteine Lowering with Folic Acid and B Vitamins in Vascular Disease. *The New England journal of medicine* 2006;1567-77. doi:10.1056/NEJMoa060900.
  29. Billot L, Sc M, Cass A, Gattas D, Glass P et al. Hydroxyethyl Starch or Saline for Fluid Resuscitation in Intensive Care. *The New England journal of medicine* 2012;1901-11. doi:10.1056/NEJMoa1209759.
  30. Olbrich H, Hausleiter J, Richardt G, Henersdorf M, Empen K, Fuernau G et al. Intraaortic Balloon Support for Myocardial Infarction with Cardiogenic Shock. *The New England journal of medicine* 2012;1287-96. doi:10.1056/NEJMoa1208410.
  31. Design S. Localized intracoronary gamma-radiation therapy to inhibit the recurrence of restenosis after stenting. *The New England journal of medicine* 2001;344(4):250-6. doi:10.1056/NEJM200101253440402.
  32. Bosch J, Gerstein HC, Dagenais GR, Díaz R, Dyal L, Jung H et al. N-3 Fatty Acids and Cardiovascular Outcomes in Patients With Dysglycemia. *The New England journal of medicine* 2012;367(4):309-18. doi:10.1056/NEJMoa1203859.
  33. Birnie DH, Healey JS, Wells G, Verma A, Tang AS, Krahn AD et al. Pacemaker or defibrillator surgery without interruption of anticoagulation. *The New England journal of medicine* 2013;368(22):2084-93. doi:10.1056/NEJMoa1302946.
  34. Harrington R, Stone GW, McNulty S, White HD, Lincoff a M, Gibson CM et al. Platelet inhibition with cangrelor in patients undergoing PCI. *The New England journal of medicine* 2009;361(24):2318-29. doi:10.1056/NEJMoa0908628.
  35. Chase AJ, Edwards RJ, Hughes LO, Berry C, Ch B et al. Randomized trial of preventive angioplasty in myocardial infarction. *The New England journal of medicine* 2013;369(12):1115-23. doi:10.1056/NEJMoa1305520.
  36. Stapf C, Delcourt C, Lindley R, Robinson T, Woodward M, et al. Rapid blood-pressure lowering in patients with acute intracerebral hemorrhage. *The New England journal of medicine* 2013;368(25):2355-65. doi:10.1056/NEJMoa1214609.
  37. Wühl E, Trivelli A, Picca S, Litwin M, Peco-Antic A, Zurowska A et al. Strict blood-pressure control and progression of renal failure in children. *The New England journal of medicine* 2009;361(17):1639-50. doi:10.1056/NEJMoa0902066.
  38. Horn J, Hovdenes J, Åneman A, Al-subaie N, Winkel P et al. Targeted temperature management at 33°C versus 36°C after cardiac arrest. *The New England journal of medicine* 2013;369(23):2197-206. doi:10.1056/NEJMoa1310519.
  39. Spertus JA, Herrin J, Lin Z. Telemonitoring in Patients with Heart Failure. *The New England journal of medicine* 2013;2301-9. doi:10.1056/NEJMoa1010029.
  40. Agus MSD, Steil GM, Wypij D, Costello JM, Laussen PC, Langer M et al. Tight glycemic control versus standard care after pediatric cardiac surgery. *The New England journal of medicine* 2012;367(13):1208-19. doi:10.1056/NEJMoa1206044.
  41. Mauriaucourt P, Bragança C, Billères X, Fuster P, Thiercelin D, Debaty G et al. Vasopressin and epinephrine vs. epinephrine alone in cardiopulmonary resuscitation. *The New England journal of medicine* 2008;359(1):21-30. doi:10.1056/NEJMoa0706873.
  42. Clemens F, Elbourne D, Darbyshire J, Pocock S. Data monitoring in randomized controlled trials: surveys of recent practice and policies. *Clin Trials* 2005;2(1):22-33.
  43. Pocock SJ. Current controversies in data monitoring for clinical trials. *Clin Trials* 2006; 3(6):513-21. doi:10.1177/1740774506073467.
  44. Lin JY, Lu Y. Establishing a data monitoring committee for clinical trials. *Shanghai Arch Psychiatry* 2014; 26(1): 54-6. doi:10.3969/j.issn.1002-0829.2014.01.009.