

# Análisis interino en ensayos clínicos: un estudio meta-epidemiológico

## Interim analysis in clinical trials: A meta-epidemiological study / Análise interino em ensaios clínicos: um estudo meta-epidemiológico

Laura Acevedo Ortiz<sup>1</sup>, Daniela Jaramillo Velásquez<sup>2</sup>, Jorge Hernando Donado Gómez<sup>2</sup>

Fecha de recibido:

03 de febrero de 2017

Fecha de aprobación:

04 de mayo de 2017

### RESUMEN

**Objetivo:** la responsabilidad ética del investigador obliga al monitoreo de la seguridad de los participantes a través del estudio y, por ende, se requiere de un comité de monitoreo de datos, cuya tarea principal es el análisis interino que se refiere a la supervisión de variables como beneficios dramáticos, efectos adversos, mortalidad y futilidad, que lleven a la terminación temprana del estudio. El objetivo de esta investigación es determinar los métodos estadísticos más utilizados en el análisis interino en los ensayos clínicos aleatorizados y publicados en 2016 en una revista médica general de alto factor de impacto.

**Metodología:** se realizó un estudio meta-epidemiológico descriptivo, constituido por ensayos clínicos aleatorizados publicados en el The New England Journal of Medicine, desde el 7 de enero hasta el 10 de noviembre de 2016.

**Resultados:** se analizaron 104 artículos con una muestra promedio de  $5\,531 \pm 762$  pacientes y un tiempo de seguimiento promedio de  $31 \pm 45$  meses. De los estudios evaluados se encontró que 88 (84.61%) realizaron análisis interino y 16 (15.38%) no lo realizaron. Los métodos estadísticos utilizados en los artículos fueron O'Brien-Fleming en 30 (34.09%), Haybittle-Peto en 13 (14.77%); otros métodos en dos (2.27%) y no especificado en 43 (48.86%) de ellos.

**Conclusiones:** la mayoría de los ensayos clínicos aleatorizados realizó análisis interino y lo describen en su protocolo. Aproximadamente la mitad de los estudios no especificó el método estadístico; sin embargo, el método descrito más utilizado fue O'Brien-Fleming y, en segundo lugar, Haybittle-Peto y en ningún estudio se especificó el método de Pocock.

**Palabras clave:** interpretación estadística de datos; ensayo clínico; comités de monitoreo de datos de ensayos clínicos.

Forma de citar este artículo:

Acevedo L, Jaramillo D, Donado JH. Análisis interino en ensayos clínicos: un estudio meta-epidemiológico. Med U.P.B. 2017;36(2):134-137. DOI:10.18566/medupb.v36n2.a05

### ABSTRACT

**Objective:** The ethical responsibility of the researcher requires that the safety of participants be monitored throughout the study. Thus, a data monitoring committee is required, whose main function is the interim analysis, which monitors variables such as dramatic benefits, adverse effects, mortality and futility that lead to the early termination of a study. The objective of this study is to determine the most frequently used statistical methods in interim analysis in randomized clinical trials published during 2016 in a high impact factor medical journal.

**Methodology:** Meta-epidemiological descriptive study, consisting of randomized clinical trials published in The New England Journal of Medicine from January 7 to November 10 of 2016.

**Results:** 104 articles were analyzed, where the mean sample was  $5,531 \pm 762$  patients and average follow-up time was  $31 \pm 45$  months. Of the studies evaluated, 88 (84.61%)

1. Facultad de Medicina, Universidad CES. Medellín, Colombia.
2. Profesor Titular Universidad Pontificia Bolivariana y Epidemiólogo Clínico Hospital Pablo Tobón Uribe.

Dirección de

correspondencia: Laura Acevedo Ortiz. Correo electrónico: lauracvdo@gmail.com

performed interim analysis, while 16 (15.38%) did not. Regarding statistical methods used, O'Brien-Fleming was used in 30 (34.09%) studies, Haybittle-Peto in 13 (14.77%), other methods in two (2.27%), and the method was not specified in 43 (48.86%).

**Conclusions:** most of the randomized clinical trials reviewed had interim analysis and described it in their protocol. Approximately half of the studies did not specify the statistical method; however, the most commonly described method was O'Brien-Fleming, followed by Haybittle-Peto; no study used the Pocock method.

**Keywords:** data interpretation, statistical; clinical trial; clinical trials data monitoring committees.

## RESUMO

**Objetivo:** a responsabilidade ética do investigador obriga à monitoração da segurança dos participantes através do estudo e, por consequência, se requiere de um comité de monitoração de dados, cuja a tarefa principal é a análise interino que se refere à supervisão de variáveis como benefícios dramáticos, efeitos adversos, mortalidade e futilidade, que levem à terminação precoce do estudo. O objetivo desta investigação é determinar os métodos estatísticos mais utilizados na análise interino nos ensaios clínicos aleatorizados e publicados em 2016 em uma revista médica geral de alto fator de impacto.

**Metodologia:** se realizou um estudo meta-epidemiológico descritivo, constituído por ensaios clínicos aleatorizados publicados no *The New England Journal of Medicine*, desde o dia 7 de Janeiro até o dia 10 de novembro de 2016.

**Resultados:** se analisaram 104 artigos com uma amostra média de  $5\,531 \pm 762$  pacientes e um tempo de seguimento médio de  $31 \pm 45$  meses. Dos estudos avaliados se encontrou que 88 (84.61%) realizaram análise interino e 16 (15.38%) não o realizaram. Os métodos estatísticos utilizados nos artigos foram *O'Brien-Fleming* em 30 (34.09%), *Haybittle-Peto* em 13 (14.77%); outros métodos em dois (2.27%) e não especificado em 43 (48.86%) deles.

**Conclusões:** a maioria dos ensaios clínicos aleatorizados realizou análise interino e o descrevem no seu protocolo. Aproximadamente a metade dos estudos não especificou o método estatístico; mas, o método descrito mais utilizado foi *O'Brien-Fleming* e, em segundo lugar, *Haybittle-Peto* e em nenhum estudo se especificou o método de *Pocock*.

**Palavras chave:** Interpretação estatística de dados; Ensaio clínico; Comitês de monitoração de dados de ensaios clínicos.

## INTRODUCCIÓN

En los ensayos clínicos aleatorizados de superioridad se rechaza una hipótesis nula en la que no existe diferencia entre la intervención y el control. Esto se logra a través del análisis de los datos acumulados a partir de cada participante del estudio; sin embargo, la responsabilidad ética del investigador obliga al monitoreo de la seguridad y el beneficio clínico de los participantes de forma secuencial a través del desarrollo del estudio. Por esta razón se requiere de un comité de monitoreo de datos para velar por estos principios, pero no todos los estudios requieren de un comité de monitoreo por cuanto ello

depende del tipo de estudio, la duración y la población. Es obligatorio el monitoreo por parte del comité en enfermedades que amenacen la vida, estudios de larga duración, población con pacientes discapacitados o pediátricos o gran sospecha de un efecto potencialmente dañino para los pacientes<sup>1-3</sup>.

El comité debe estar conformado por un grupo independiente de los investigadores, que no debe estar cegado en cuanto a las intervenciones usadas para aumentar la credibilidad. Debe ser un equipo multidisciplinario de especialistas en el tema de estudio: epidemiólogos experimentados en análisis/diseño y expertos en bioética. Los miembros del comité deben tener conoci-

miento de la metodología de los ensayos clínicos, incluidos métodos estadísticos para calcular el efecto del azar e imparcialidad financiera, intelectual y emocional<sup>1,2,4</sup>.

La principal tarea del comité de monitoreo de datos es el análisis interino que consiste en la supervisión de variables que lleven a la terminación temprana del estudio. Los principales aspectos para detectar son los beneficios dramáticos, los efectos adversos, la mortalidad y la futilidad, que se realiza por dos tipos de directrices estadísticas, una mediante el intervalo de confianza IC 95% y otra con el poder condicional<sup>1,2,5,6</sup>.

Adicionalmente, se deben analizar otras variables en el estudio como seguridad, integridad, aspectos del diseño, progreso y económicos. El comité revisa el protocolo, los procedimientos de prueba y las reglas de detención, el número de sesiones y el plan de análisis. Con base en estas revisiones el investigador tendrá recomendaciones acerca de las modificaciones, continuación o terminación del estudio. Cabe resaltar que las decisiones finales son responsabilidad del patrocinador o comité ejecutivo<sup>1-3,5,6</sup>.

El análisis interino se puede obtener a través de diferentes métodos estadísticos como los secuenciales clásicos y métodos secuenciales agrupados entre los que se encuentran: Pocock, O'Brien y Fleming y Haybittle-Peto<sup>2,6</sup>.

El objetivo de este estudio es determinar los métodos estadísticos más utilizados en el análisis interino en los ensayos clínicos aleatorizados y publicados durante el 2016 en una revista médica general de alto factor de impacto.

## METODOLOGÍA

Se realizó un estudio meta-epidemiológico descriptivo, constituido por ensayos clínicos aleatorizados y publicados en el *The New England Journal of Medicine*, desde el 7 de enero hasta el 10 de noviembre de 2016. Dos investigadores se encargaron de la búsqueda de los artículos y el análisis de cada uno para determinar su elegibilidad. Se incluyeron los ensayos clínicos aleatorizados de grupos paralelos que tuvieran protocolo disponible y se excluyeron los ensayos clínicos con diseño *crossover*. No hubo limitación de idioma.

Después de incluir los artículos se obtuvo el texto completo y los suplementos del protocolo, se analizaron y se extrajeron las siguientes variables: área de estudio, número total de muestra, tiempo promedio de seguimiento, realización o no de análisis interino, método estadístico utilizado, número de sesiones realizadas, variable por la cual se determinó el número de sesiones: por eventos, muestra o tiempo. Si se encontraban diferentes intervalos de tiempo para las sesiones se escogía el intervalo más corto y si no se describía alguna de las variables se clasificaba como no especificado.

El análisis de los datos se realizó para las variables cuantitativas con medidas de tendencia central (como media y mediana), medidas de dispersión como desviación estándar, rango y rango intercuartil (RIC). Para las variables cualitativas se usaron frecuencias absolutas y relativas. Se estimó la concordancia de la recolección de la información entre los investigadores con el coeficiente kappa. Para comparar las medias se utilizó la t de student de muestras independientes. Se empleó el paquete estadístico Epidat versión 4.2.

El estudio fue conducido de acuerdo con los principios de la Declaración de Helsinki, versión 2013.

## RESULTADOS

Se identificaron 106 artículos que cumplían los criterios de inclusión. Se excluyeron dos por tener diseño *crossover* y quedaron 104 artículos para su respectivo análisis. Se encontró que 55 (52.88%) pertenecían al área de medicina interna, 16 (15.38%) a cirugía, 12 (11.53%) a obstetricia, ocho (7.69%) a pediatría, y 13 (12.5%) a otras áreas. Los ensayos clínicos seleccionados tenían una muestra promedio de  $5\,531 \pm 762$  pacientes, una mediana de 766 (RIC 329.75 a 2 423.25) y un rango de 66 a 204 700 pacientes. El tiempo de seguimiento promedio fue de  $31 \pm 45$  meses, con una mediana de 24 (RIC 6.75 a 44.7) y un rango de uno a 120 meses.

De los estudios evaluados se encontró que 88 (84.61%) realizaron análisis interino y 16 (15.38%) no lo realizaron. Los métodos estadísticos usados fueron O'Brien-Fleming en 30 (34.09%), Haybittle-Peto en 13 (14.77%), otros métodos en dos (2.27%) y no fue especificado en 43 (48.86%) de ellos.

No se encontraron diferencias entre los estudios que hicieron análisis interino y los que no, con respecto al tamaño de la muestra  $2\,934 \pm 3\,911$  vs  $5\,906 \pm 2\,6113$  con una diferencia de  $-2\,972$  (IC 95%  $-16\,000$  a  $10\,053$ ). Tampoco se evidenció diferencia en el tiempo de seguimiento,  $27.05 \pm 27.5$  vs  $32.67 \pm 31.16$  con una diferencia  $-5.6$  (IC 95%  $-22.1$  a  $10.9$ ).

La media del número de sesiones fue de  $2.19 \pm 1.15$  y la mediana fue de 2 (RIC 1-3), rango de uno a seis sesiones. Sin embargo, 32 (30.76%) artículos no especificaron el número de sesiones. El número de sesiones de análisis interinos se determinó en 22 (25%) por muestra, en 17 (19.31%) por tiempo, en 27 (30.68%) por evento y en 27 (30.68%) no fue especificado. En las sesiones determinadas por tiempo se encontró una media de  $9.41 \pm 5.07$  meses, una mediana de nueve (RIC 6 a 12) y un rango de uno a 24 meses.

El coeficiente de kappa entre los evaluadores fue de 0.81 (IC 95% 0.68 a 0.94).

## DISCUSIÓN

Se encontró que la mayoría de los ensayos clínicos aleatorizados publicados en el 2016 en el *The New England Journal of Medicine* realizó análisis interino y lo describieron en su protocolo. Aproximadamente, la mitad de los estudios no especificó el método estadístico; sin embargo, el método descrito más usado fue O'Brien-Fleming, en segundo lugar, Haybittle-Peto y en ningún estudio se especificó el método de Pocock.

Los métodos secuenciales clásicos están diseñados para reducir el número de participantes que deben entrar a un estudio para realizar una regla estadística después de cada sujeto que sea ingresado y haya recibido la intervención o el control. Se establecen límites de superioridad, inferioridad y equivalencia, se detiene el estudio si alcanza el límite superior o inferior. Generalmente, los métodos clásicos son de sesión abierta, es decir, no se sabe en qué momento se detendrá el análisis porque no se fija un número de monitoreos<sup>2,3</sup>.

En los métodos secuenciales agrupados se usan sesiones cerradas en las que se establece previamente el nivel de significancia y el poder ( $\alpha$  y  $1-\beta$ ) y se hacen pruebas repetidas a dos colas. El método de Pocock divide la muestra total en grupos con igual número de participantes y establece un número de pruebas, fija un valor de significancia antes de iniciar los análisis y aplica el mismo valor a cada grupo. Cuando el valor Z es mayor que  $\alpha$  se detiene el estudio y es más fácil en etapas tempranas que en etapas tardías. El método Haybittle-Peto es más conservador porque pone un valor muy bajo de significancia y hace difícil detener el estudio en etapas tempranas. El método de O'Brien y Fleming crea un  $\alpha$  corregido para cada grupo y crea un problema de pruebas múltiples, es decir, a mayor número de análisis interinos aumenta la probabilidad de obtener un resultado estadísticamente significativo por azar (concepto llamado gasto de la "p" o de la probabilidad de error tipo 1), lo que hace complejo detener el estudio en etapas tempranas y aumenta la posibilidad en etapas tardías. Para el ajuste de este error se han creado métodos como Lan y DeMets basados en la "función de gasto", que permiten tener pruebas estadísticas más flexibles, en las que los análisis interinos no están prefijados, sino que se determinan según el porcentaje de información que se va recogiendo como el ritmo de inclusión de los pacientes, tiempo y número de eventos<sup>2,3,6,9</sup>.

Dentro de los protocolos de los ensayos clínicos se debe especificar la composición del comité de monitoreo de datos, estructura de reporte que tendrá, independencia respecto al patrocinador y ausencia de intereses económicos. Si todavía no está conformado se debe exponer el tamaño y las características de los participantes que se

tienen estimados y, si no se conforma un comité, se debe dar una explicación de porqué no es necesario para ese estudio. Si esta información no se encuentra detallada en el protocolo se debe indicar en qué parte se puede encontrar. También se debe hacer una descripción del análisis interino con las reglas de detención que se van a utilizar y cuando se va a realizar cada uno de los análisis (tiempo o indicaciones específicas como número de muestra o eventos)<sup>9</sup>.

En estudios previos se ha descrito la proporción de estudios que han realizado análisis interinos y se ha encontrado que el 71% (106/150) de los protocolos de los ensayos clínicos de cáncer con desenlace tiempo a evento en Italia, en el 2005 lo realizó. Otro estudio reportó que entre 1994 y 1995 el 19% (13/70) de los ensayos clínicos aleatorizados aprobados en Dinamarca realizó análisis interino. Más recientemente se encontró que aproximadamente el 40% de los estudios registrados en ClinicalTrial.gov del 2007 al 2010 reportó tener un comité de monitoreo de datos. De los 86 protocolos de los ensayos clínicos de cáncer con desenlace tiempo a evento que tenían análisis interino, 100% describió el tiempo en el que se realizaron los análisis, 91% especificó las reglas de detención y 94% describió el método estadístico utilizado<sup>9</sup>. En el presente estudio se halló que el 85% tenía análisis interino y no hubo diferencia ni por tamaño de muestra ni por tiempo de seguimiento con respecto a los que no lo hicieron.

De los estudios analizados la mayoría realizó entre uno a tres sesiones y, al evaluar las variables por las cuales se determinaba el número de sesiones, se encontró que podía ser por tiempo, evento, muestra o no especificado. Se encontró que los no especificados y por evento tuvieron igual proporción (30.68%), seguidos de muestra y, en último lugar, por tiempo, pero ninguna de estas variables predominó de una manera significativa. Cuando las sesiones eran determinadas por tiempo se observó que la mayoría tenía un intervalo entre seis a 12 meses.

Al comparar los resultados obtenidos en este estudio con los realizados previamente se infiere que, a medida que han pasado los años, los investigadores se han concientizado más acerca de la importancia del análisis interino y se ha aumentado el número de estudios que lo incluyen. Sin embargo, casi un tercio de los artículos encontrados no especificaron el número de sesiones ni la variable por la cual se iba a determinar el número de sesiones y casi la mitad no especificó el método estadístico usado y, por tanto, no se adhieren a las recomendaciones que debe cumplir un protocolo de un ensayo clínico aleatorizado<sup>9</sup>.

Hay que tener en cuenta que las limitaciones del estudio fueron incluir artículos de una sola revista y

utilizar un rango estrecho de fechas para seleccionarlos. En conclusión, aunque ha aumentado el número de estudios que realizan análisis interinos no se cumple con la recomendación para describirlo en el protocolo.

## DECLARACIÓN DE CONFLICTO DE INTERESES

Los autores declaran no tener ningún conflicto de intereses.

## REFERENCIAS

1. Committee for medicinal products for human use (CHMP). Guideline on data monitoring committees. *Statistics in Medicine* 2006; 25(10):1639-1645.
2. Friedman L, Furberg C, DeMets D, Reboussin D, Granger C. Monitoring committee structure and function and statistical methods used in interim monitoring. *Fundamentals of Clinical Trials, Fifth Edition*. Springer Cham Heidelberg New York Dordrecht London; 2015. p. 343-401.
3. Salim D. Interim analysis. *Evidence-based Obstetrics and Gynecology* 2000; 2(2):2-3.
4. Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, et al. Issues in data monitoring and interim analysis of trials. *Health Technol Assess* 2005; 9(7):1-238.
5. Thom E, Klebanoff M. Issues in clinical trial design: Stopping a trial early and the large and simple trial. *American Journal of Obstetrics and Gynecology* 2005; 193(3): 619-662
6. Pocock SJ, Clayton TC, Gregg W. Stone challenging issues in clinical trial design. *J Am Coll Cardiol* 2015; 66(25):2886-2898.
7. Muñoz SR, Bangdiwala SI. Análisis interino en ensayos clínicos: una guía metodológica. *Rev Méd Chile* 2000; 128(8):935-941.
8. Candia R, Rada G. Detención precoz de estudios clínicos randomizados: ¿beneficio para los pacientes o para el investigador?: sobrestimación de resultados, reglas de detención y comités de monitorización de datos. *Ars Medica Revista de Ciencias Médicas* 2015; 39(1):9.
9. Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ* 2013; 346:e7586.