

Interrupción precoz de los ensayos clínicos. ¿Demasiado bueno para ser cierto?

J. LATOUR-PÉREZ^a Y J.B. CABELLO-LÓPEZ^b

^aServicio de Medicina Intensiva. Hospital General Universitario de Elche. Elche. España.

^bServicio de Cardiología. Hospital Universitario de Alicante. Alicante. España. Coordinación General de CASP-España.

La interrupción de un ensayo clínico aleatorizado cuando existe una clara superioridad del tratamiento experimental constituye una exigencia ética. Sin embargo, la realización de múltiples análisis estadísticos intermedios y la interrupción precoz del ensayo aumentan la posibilidad de error alfa, favorecen la sobreestimación del efecto y facilitan la introducción de tratamientos ineficaces.

Actualmente se dispone de evidencia empírica que sugiere que los ensayos clínicos interrumpidos precozmente sobreestiman el efecto de los nuevos tratamientos y frecuentemente adolecen de una insuficiente especificación de las razones que justifican la interrupción del ensayo.

Todo ello subraya la necesidad de unos comités de monitorización de los datos verdaderamente independientes, y aconsejan la utilización de reglas restrictivas predefinidas para la interrupción del ensayo. Por su parte, los clínicos deberían proveerse de la adecuada dosis de escepticismo y de habilidades de lectura crítica para explorar algunos aspectos clave de estos estudios.

PALABRAS CLAVE: ensayos clínicos aleatorizados, comités de monitorización de los datos, análisis intermedios, error estadístico, ética.

EARLY INTERRUPTION OF CLINICAL TRIALS. TOO GOOD TO BE TRUE?

The interruption of a randomized clinical trial when the experimental treatment is clearly superior is an ethical obligation. However, performing multiple interim statistical analyses and the early interruption of the trial increase the possibility of alpha error, favor the overestimation of the effect, and facilitate the introduction of ineffective therapies data monitoring.

Currently available empirical evidence suggests that early interrupted clinical trials overestimate the effects of new treatments and that the reasons used to justify interrupting the trial are often not sufficiently specified.

This underlines the need for truly independent data monitoring committees and recommends the use of predefined restrictive rules for the interruption of a trial. Furthermore, clinicians should provide themselves with an adequate dose of skepticism and critical reading skills to enable them to explore some key aspects of these studies.

KEY WORDS: randomized clinical trials, data monitoring committees, intermediate analyses, statistical error, ethics.

Con frecuencia, durante la realización de un ensayo clínico aleatorizado es preciso monitorizar los desenlaces clínicos mediante la realización de sucesivos análisis estadísticos precoces (*interim analysis*). El principal objetivo de estos análisis intermedios es detectar precozmente aquellas situaciones en las que uno de los brazos del estudio (el grupo experimental o el grupo control) es claramente superior al otro, al

Correspondencia: Dr. J. Latour Pérez.
UCI. Hospital General Universitario de Elche.
Camí Vell de l'Almàssera, 11.
03203 Elche. Alicante.
Correo electrónico: jlatour@coma.es

Manuscrito aceptado el 4-VIII-2006.

objeto de interrumpir precozmente el ensayo y así evitar el sometimiento de seres humanos al tratamiento menos efectivo o seguro. Este proceder constituye una exigencia ética cuando existen pruebas claras de efectos adversos excesivos, evidencia clara de beneficio o una alta probabilidad de que el tratamiento sea fútil. Sin embargo, la interrupción precoz del ensayo puede plantear serios problemas, tanto metodológicos como éticos.

En primer lugar, la realización de análisis intermedios mediante el abordaje estadístico habitual genera un problema de multiplicidad que aumenta la probabilidad de error α . En efecto, si se realizan varios análisis intermedios y el investigador examina los datos acumulados con una $p < 0,05$ en cada uno de ellos, el error α global aumenta a 0,08 tras dos análisis, a 0,11 tras tres análisis y a 0,19 tras 10 análisis¹. Se impone, por tanto, la exigencia de usar reglas estadísticas ajustadas que tomen en cuenta esta multiplicidad de análisis.

En segundo lugar, la interrupción precoz del ensayo (sea o no justificada) favorece la sobrestimación del efecto. Ello se comprende fácilmente si imaginamos un ensayo bien diseñado y ejecutado, que se repite varias veces. Debido al error aleatorio inherente al muestreo, el riesgo relativo detectado en los sucesivos ensayos oscilará alrededor de su verdadero valor. En estas condiciones, las fluctuaciones aleatorias que tienden a exagerar el efecto del tratamiento conducirán a una interrupción precoz del ensayo más fácilmente que las fluctuaciones que tienden a subestimar el efecto. Por lo tanto, los efectos estimados a partir de ensayos interrumpidos precozmente son propensos a la exageración. Dicho de otra forma, si se interrumpe el ensayo por el simple hallazgo de una p significativa, es muy probable que el efecto detectado sea un máximo aleatorio (*random high*) y que, si se continuara el ensayo, el efecto final detectado sería menor (regresión a la media).

En tercer lugar, en la medida en que favorece el riesgo de introducir tratamientos ineficaces, la interrupción precoz del ensayo plantea un nuevo dilema ético: la necesidad de no privar a los sujetos del ensayo de un tratamiento probablemente beneficioso entra en conflicto con la necesidad de proteger a futuros enfermos de la administración de un tratamiento fútil y de sus efectos adversos asociados².

Recientemente Montori et al³ han realizado una revisión sistemática que aporta importante evidencia empírica sobre los ensayos aleatorizados interrumpidos precozmente por efecto beneficioso. Estos autores detectan un aumento progresivo del número de estos ensayos publicados en los últimos años, tanto en valores absolutos (esperable por el aumento del número total de ensayos) como en términos relativos (porcentaje de ensayos clínicos indexados en Medline). Más curioso aún es que la proporción de estos ensayos publicados en las 10 revistas de mayor impacto crece más rápidamente que la proporción en el resto de las revistas, lo que sugiere que la interrupción precoz podría estar siendo considerada por la comunidad científica como un valor positivo.

En el estudio de Montori et al³, los ensayos se interrumpieron cuando se había recogido como promedio el 64% de la muestra prevista, y con un riesgo relativo estimado en conjunto de 0,53 (reducción relativa del riesgo del 47%). Además se encontró una fuerte asociación inversa entre el número de eventos y la magnitud del efecto estimada, coherente con la hipótesis de que el efecto detectado es un máximo aleatorio que sobreestima el verdadero efecto del tratamiento.

Asimismo, los autores destacan una baja especificación de las razones que condujeron a la interrupción del estudio: sólo 67 (47%) de los 143 ensayos comunicaban el tamaño de la muestra planeado, el análisis intermedio en el que se tomó la decisión de interrumpir el ensayo, y la regla que se utilizó para informar esta decisión (aspectos, todos ellos, incluidos en CONSORT)⁴. Finalmente, sólo 8 ensayos comunicaron estos tres elementos junto con las estimaciones ajustadas para el análisis intermedio.

El estudio de Montori et al³ ofrece mensajes importantes para las estructuras involucradas en la producción y publicación de los ensayos (investigadores, financiadores, comités éticos, agencias reguladoras, editores, revisores, panelistas, etc.) y, desde luego, para los clínicos empeñados en la lectura crítica de la literatura científica y enfrentados a la tarea de elegir y aplicar el mejor tratamiento para su paciente.

Para los primeros, estos resultados subrayan la importancia de un comité de monitorización del ensayo (*Data Monitoring Committee, DMC*) verdaderamente independiente^{5,6}, con derecho exclusivo de acceso a los datos tabulados por tratamiento, y con capacidad para decidir cuándo existe suficiente evidencia para interrumpir el ensayo y cuándo éste debe proseguir.

El trabajo de este comité puede verse facilitado notablemente por la implantación de adecuados procedimientos de trabajo⁵, y específicamente por el uso de «reglas de decisión» predefinidas para la interrupción del ensayo. De hecho, existen diversas reglas de interrupción de un ensayo⁷, como la llamada «regla de Pocock», la de Peto o la de O'Brien-Fleming. Aunque algunas son más restrictivas que otras, todas ellas exigen un nivel de probabilidad mucho más exigente que el habitual para interrumpir un ensayo clínico. Otro método que ha ganado popularidad como criterio para decidir la interrupción precoz del ensayo es el método de DeMets⁸ (función del gasto de alfa, *alpha spending function*), que permite la realización de análisis intermedios no planificados previamente.

Para los lectores de la literatura médica el mensaje es que ante el supuesto «valor» de los resultados de los ensayos interrumpidos precozmente deberían proveerse de la adecuada dosis de escepticismo, y de habilidades para explorar algunos aspectos clave de estos ensayos.

Por una parte es preciso prestar atención al número de análisis estadísticos realizados y al método utilizado para prevenir el problema de los resultados falsos positivos. En ocasiones esto no es sencillo, ya que el número de análisis no siempre se comunica de forma explícita. Una clave podría ser que el valor comu-

nicado de p sea ligeramente menor de 0,05, lo que puede significar que los investigadores han hecho análisis repetidos hasta alcanzar esa p significativa.

Por otra parte, es importante examinar si el criterio de interrupción ha tenido en cuenta el número de análisis intermedios realizados. Por ejemplo, cuando se prevén dos análisis intermedios, para interrumpir el ensayo con la regla de Peto se exige una $p < 0,001$, mucho menor que el punto de corte habitualmente consignado en muchos ensayos interrumpidos precozmente.

Asimismo, ante un estudio interrumpido precozmente por beneficio, debería compararse el tamaño muestral inicialmente previsto con el incluido finalmente en el ensayo. Una desproporción entre los dos, o un riesgo relativo extremadamente bajo, puede alertar al lector sobre la posibilidad de una sobreestimación del efecto, especialmente si el número de eventos o desenlaces es bajo.

En los últimos años se han publicado diversos ensayos ilustrativos. Por ejemplo, el estudio de Amato et al⁹ sobre la ventilación con volúmenes bajos, reúne las características de ser un ensayo con múltiples análisis intermedios, un número relativamente bajo de eventos clínicos y un efecto extremadamente beneficioso (reducción de la mortalidad ajustada del 81%); todo ello aconsejaba una actitud prudente y esperar a nuevos estudios que, de hecho, han moderado las expectativas iniciales^{10,11}. Por el contrario, en el estudio de la insulino terapia en pacientes críticos quirúrgicos, Van den Berghe et al¹² generaron una estimación del efecto ajustada para los múltiples análisis intermedios (riesgo relativo 0,68, IC 95%: 0,45-0,98) mucho más creíble que la estimación basada en un análisis convencional (riesgo relativo 0,58; IC 95%: 0,38-0,78). Otro ejemplo llamativo es el reciente ensayo OPTIMIST¹³ sobre el efecto de tifacogin en pacientes con sepsis grave y un *international normalized ratio* (INR) alto, en el que se realizó de forma planificada un análisis intermedio con los primeros 722 pacientes reclutados, encontrándose una reducción significativa de la mortalidad en el grupo experimental (29,1% frente al 38,9%; $p = 0,006$), que desapareció en el análisis final tras reclutar a 1.754 pacientes (34,2% frente a 33,9%; $p = 0,88$).

En síntesis, la evidencia empírica disponible muestra que los ensayos aleatorizados interrumpidos precozmente carecen frecuentemente de una descripción adecuada de los métodos utilizados para decidir la interrupción, y tienden a sobreestimar el beneficio del tratamiento, especialmente cuando el número de eventos detectados es limitado. Por tanto, lejos de

considerar la interrupción precoz del ensayo como un rasgo positivo, el lector debería abordar estos estudios con escepticismo, y los grupos comprometidos con el desarrollo de herramientas de ayuda para la realización de lectura crítica^{14,15} deberían adaptar sus instrumentos a esta circunstancia cada vez más frecuente.

Declaración de conflicto de intereses

Los autores han declarado no tener ningún conflicto de intereses.

BIBLIOGRAFÍA

- Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet*. 2005;365:1657-61.
- Pocock SJ, Hughes MD. Practical problems in interim analyses, with particular regard to estimation. *Control Clin Trials*. 1989;10 Suppl 4:209S-21S.
- Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294:2203-9.
- Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001;285:1987-91.
- Damocles Study Group. A proposed charter for clinical trial data monitoring committees: helping them to do their job well. *Lancet*. 2005;365:711-22.
- Lexchin J, Light DW. Commercial influence and the content of medical journals. *BMJ*. 2006;332:1444-7.
- Pocock SJ. When (not) to stop a clinical trial for benefit. *JAMA*. 2005;294:2228-30.
- DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. *Stat Med*. 1994;13:1341-52.
- Amato MB, Barbas CS, Medeiros DM, Magaldi RB, Schettino GP, Lorenzi-Filho G, et al. Effect of a protective-ventilation strategy on mortality in the acute respiratory distress syndrome. *N Engl J Med*. 1998;338:347-54.
- Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome Network. *N Engl J Med*. 2000;342:1301-8.
- Villar J, Kacmarek RM, Pérez-Méndez L, Aguirre-Jaime A. A high positive end-expiratory pressure, low tidal volume ventilatory strategy improves outcome in persistent acute respiratory distress syndrome: a randomized, controlled trial. *Crit Care Med*. 2006;34:1311-8.
- Van del Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schetz M, et al. Intensive insulin therapy in the critically ill patients. *N Engl J Med*. 2001;345:1359-67.
- Abraham E, Reinhart K, Opal S, Demeyer I, Doig C, Rodríguez AL, et al. Efficacy and safety of tifacogin (recombinant tissue factor pathway inhibitor) in severe sepsis: a randomized controlled trial. *JAMA*. 2003;290:238-47.
- CASP-España. Herramientas para la lectura crítica. CASP-España. 13-6-2006. 4-7-2006.
- Jaeschke R, Guyatt G, Barratt A, Walter S, Cook D, McAlister F, et al. User's guides to the medical literature. A manual for evidence-based clinical practice. Chicago: AMA Press; 2002.