

# Interrupción temprana de estudios de investigación por beneficio ¿Una trampa?

## Parte 9

How to interpret an early stopped randomized controlled study

Botaro F<sup>(1)(2)(3)</sup>

<sup>(1)</sup>Jefe de Emergencias Hospital Británico

<sup>(2)</sup>Servicio de Clínica Médica Hospital Británico

<sup>(3)</sup>Coordinador Comité de Revisión Institucional Hospital Británico

febottaro@hotmail.com

Fecha recepción: 1/11/2016

Fecha aprobación: 5/12/2016



CÓMO SE LEE

HEMATOLOGÍA  
Volumen 20 n° 3: 358 - 362  
Septiembre - Diciembre 2016

**Palabras clave:** interrupción temprana, sesgo, sobre-estimación.

**Keywords:** early interruption, stopped, overestimation.

*“Whatever you are, be a good one”.*  
Abraham Lincoln

Uno de los puntos centrales en la interpretación o revisión crítica de ensayos clínicos aleatorizados es determinar si los hallazgos del estudio son válidos; esto significa, valorar si cumple con una serie de criterios que aseguren que la posibilidad de sesgo en el mismo haya sido mínima<sup>(1)</sup>.

Uno de los ítems a valorar en la estimación de riesgo de sesgo de los ensayos clínicos aleatorizados es si el estudio fue detenido tempranamente (o prematuramente) antes de tiempo, es decir antes de que se hubieran reclutado la totalidad de los pacientes predefinida por los investigadores.

En esta entrega revisaremos los potenciales problemas que pueden surgir cuando se da esta situación, y por qué es importante siempre tenerlo en cuenta.

### **Cálculo del tamaño de la muestra (o población del estudio)**

Al desarrollar un protocolo de investigación de un ensayo clínico controlado aleatorizado (ECCA), los investigadores deben realizar el cálculo del tamaño muestral que permita reclutar la suficiente cantidad de pacientes para refutar o validar las hipótesis planteadas. Es decir, que permita demostrar las diferencias esperadas entre el grupo intervención (el grupo que ha recibido la terapéutica en experimentación) y el grupo control (quienes no la han recibido). Este paso es de suma importancia para evitar que puedan ocurrir dos situaciones:

1. que se hallen diferencias entre los grupos (intervención y control) pero que éstas sean debidas

al azar, y no a una real diferencia en la eficacia de la intervención sobre el grupo control (se denomina error tipo I o alfa);

2. que no se halle diferencia alguna entre los grupos cuando ésta en realidad sí existe (se denomina error tipo II o error beta).

Por convención, para el cálculo de tamaño muestral se asume habitualmente un error alfa aceptable de 5% (0,05) y un error beta de 20% (0,2, es decir un poder del estudio del 80%).

La capacidad de un estudio para hallar diferencias entre dos grupos, cuando esta diferencia existe en realidad, se denomina poder de un estudio. Cuanto mayor es el tamaño muestral de un estudio (cantidad de pacientes) y/o mayor el número de eventos, mayor será el poder del estudio para detectar diferencias entre los grupos cuando realmente las haya. Sin embargo, como los recursos para investigación son limitados, no es posible reclutar la “mayor cantidad” de sujetos en un estudio, sino la cantidad suficiente de sujetos que permitan hallar las diferencias estimadas (o predefinidas) por los investigadores en su protocolo.

Como ya podemos ver, un ECCA que debió ser interrumpido antes de tiempo por cuestiones de “poder” estadístico puede no mostrar diferencias reales entre los grupos cuando éstas realmente existen. En esas situaciones, los investigadores deberán reportar el poder del estudio con los sujetos incluidos.

### **Interrupción temprana de un ECCA**

Puede haber varios motivos por los que los investigadores de un ECCA decidan interrumpirlo antes de concluir el reclutamiento de sujetos planeado:

- a) por presunción de que la intervención en estudio puede ser perjudicial para los pacientes: esto puede ocurrir luego de observar resultados negativos en un análisis interino;
- b) porque ha surgido un nuevo estándar de cuidado que hace que alguna de las ramas del estudio ahora sea no ética, obsoleta o no apta de ser administrada a los sujetos de estudio;
- c) por la aparición de eventos adversos serios que hagan pensar a los investigadores que los mismos no justifiquen la administración del tratamiento en estudio;
- d) incapacidad de los investigadores de reclutar un número adecuado de sujetos, o el número

de sujetos necesarios de acuerdo al cálculo de tamaño muestral. Esta situación es la única en que el análisis de los resultados sobre los sujetos reclutados hasta ese momento tendría validez; ya que esta muestra, si bien más pequeña, puede ser asumida como una aproximación estimada de lo que hubiese pasado si se hubiese reclutado la totalidad de los sujetos. Claro está que afectaría y disminuiría el poder del estudio (es decir, la capacidad de hallar diferencias entre dos grupos cuando esta diferencia realmente existe). Por lo que si no se hallaran diferencias, esto podría deberse a problemas de poder de la muestra y no debido a que las ramas del estudio no sean diferentes;

- e) sin embargo, la razón con mayor impacto en la práctica clínica de interrupción temprana de un estudio es que los investigadores hayan “notado”, en un análisis interino de los datos del estudio, un efecto favorable de la rama intervención o tratamiento experimental que interpreten que hay pocas posibilidades que sea por azar. Habitualmente en estas situaciones, como el número de pacientes reclutados es bajo, el tamaño del efecto observado suele ser muy grande o importante.

### **Interrupción temprana de un ECCA por “beneficio”**

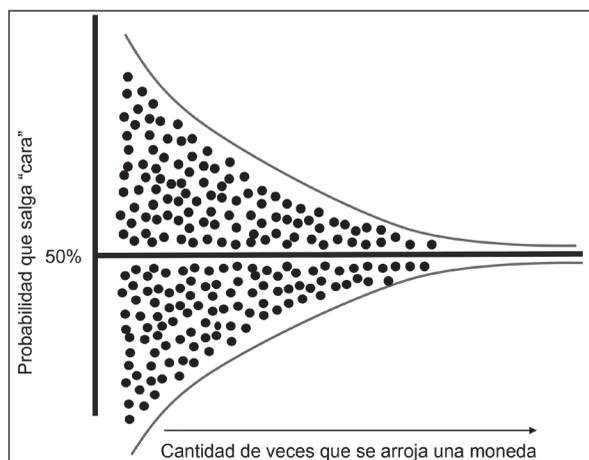
Esta situación se da a menudo en la práctica clínica, y habitualmente estos estudios reciben mucha atención y difusión, ya que los beneficios observados tienden a ser grandes. Suelen ser publicados generalmente en revistas de primer orden y en múltiples ocasiones sus resultados prevalecen por sobre los de otros estudios que pueden evidenciar efectos opuestos. Una revisión sistemática en 2005 halló que 92 sobre 143 ECCAs interrumpidos tempranamente fueron publicados en revistas de alto impacto<sup>(2)</sup>. Los estudios del área hematológica no son ajenos a esta realidad, y esta especialidad (especialmente en onco-hematología) contribuye al 20% de los estudios detenidos antes de tiempo<sup>(2)</sup>. Llamativamente, la publicación de un ECCA detenido temprano por beneficio genera habitualmente la publicación en el corto plazo inmediato de otros estudios que evalúan la misma pregunta de investigación, indicando que otros investigadores han sido más escépticos sobre interrumpir sus estudios en forma temprana<sup>(3)</sup>.

La decisión de suspender un ECCA antes de finali-

zar el reclutamiento de la totalidad planeada de pacientes si bien puede favorecer que un tratamiento eficaz pueda estar disponible para un gran grupo de personas, también puede favorecer que asumamos eficacia por cuestiones de azar y eventualmente dañemos o administremos intervenciones fútiles a este grupo de pacientes.

**¿Por qué un ECCA interrumpido antes de su plan de reclutamiento por beneficio observado en un análisis interino puede llevarnos asumir efectos que no son reales?**

En líneas generales, la mayoría de las veces que se da esta situación existe una sobre-estimación sobre el verdadero tamaño o magnitud del efecto de la intervención que no es real, sino por azar. Esto se debe a que en las etapas iniciales de un ECCA, existe un riesgo alto, tanto de subestimar como de sobestimar el efecto de una intervención<sup>(4)</sup>. En muchas ocasiones esta sobrestimación puede ser de gran magnitud, y cuantos más análisis interinos (momentos en los que se analizan los resultados del estudio antes de la finalización del mismo con los sujetos incluidos hasta ese momento) más chances de hallar efectos, o tamaños de efectos, irreales. A medida que progresa el estudio, al incluir cada vez más sujetos y, por lo tanto, aumentar el número de eventos, esta posibilidad se va reduciendo. En forma análoga, ocurre algo similar si tiráramos múltiples veces una moneda al aire (ver **Figura 1**).

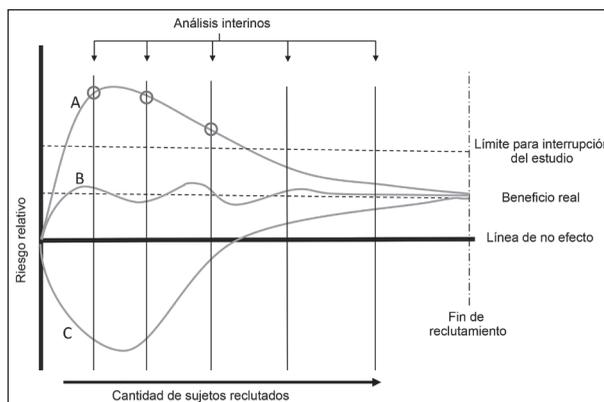


**Figura 1.** Distribución de probabilidades de que salga “cara” al tirar una moneda. Luego de arrojarla una cantidad importante de veces las probabilidades de que salga “cara” o “ceca” se irán aproximando cada vez más al 50%.

Al principio, cuando el número de veces que se arrojó la moneda es pequeño, las probabilidades

de que salgan “cara o ceca” serán muy variables y extremas, pudiendo oscilar fácilmente entre los extremos del arco de posibilidades (0 a 100%). Al ir incrementando el número de veces que se arroja la moneda las proximidades irán de forma paulatina tendiendo a acercarse a 50%<sup>(5)</sup>.

En forma similar ocurre en un ECCA cuando el tamaño muestral y el número de eventos son pequeños, por lo que las posibilidades de obtener “efectos” extremos irreales es más alta. A medida que mejora el reclutamiento y aumenta el número de eventos del estudio, el tamaño del efecto observado oscilará probabilísticamente cada vez menos. En la **Figura 2** (adaptada de Guyatt y col.<sup>(4)</sup>) podemos observar tres curvas que grafican a modo de ejemplo las oscilaciones posibles de los resultados “transitorios” de un ECCA.



**Figura 2.** Reproducido y modificado de Bassler, D., Montori, V., Devereaux, P y cols. (2015) *Randomized trials stopped early for benefit*. In: Guyatt G, Drummond R, Meade MO, Cook DJ, editors. *User’s Guide To The Medical Literature. A Manual for Evidence-Based Clinical Practice*. JAMA & Archives Journals, American Medical Association; 2015. (3rd edition, pp. 134-141).

Al llegar al final del reclutamiento las tres curvas convergen proporcionando una estimación de efecto similar, en este caso claramente diferente de la línea de no efecto. Sin embargo, “en el camino” las tres curvas poseen oscilaciones por azar diferentes. En los casos de las curvas B y C no se verá afectada la decisión de los investigadores de terminación temprana del estudio, ya que el efecto observado en los análisis interinos no ha sobrepasado el límite de “beneficio mayor” que justifica la interrupción del estudio. Sin embargo, la curva A, en los tres primeros análisis interinos, presenta riesgos relativos superlativos (sobre-estimados) que están por arriba del

umbral que los investigadores pueden pautar como justificación de terminación temprana del estudio.

### ¿Cómo se decide detener o interrumpir un ECCA antes de tiempo?

No existe un acuerdo general o regla universalmente aplicada para aplicar. La decisión deberá cotejarse en relación a la patología en estudio, intervención y las consecuencias de la detención. Los investigadores estarán en un dilema ético importante: interrumpir el estudio proclamando beneficio para no demorar la oferta de la intervención, versus el riesgo de que la interrupción genere una sobrestimación irreal del efecto.

No hay una conducta definida y muchos autores recomiendan simplemente no realizar análisis interino hasta no finalizar el reclutamiento del estudio. Claro está, que si bien esta acción puede evitar que lleguemos a conclusiones de beneficio irreales, también podría exponer a un grupo de sujetos a riesgo de eventos adversos innecesarios si la intervención fuera fútil o dañina. Otros autores recomiendan una solución intermedia, estableciendo criterios pre-especificados para la interrupción prematura del estudio adecuadamente definidos de acuerdo al número de eventos. Esto implica limitar y definir a priori el número de análisis interinos, definir un comité de análisis de datos independiente de los investigadores para la recomendación de interrupción temprana (esto debe estar explicitado en la publicación), establecer un número mínimo de eventos para la realización del primer análisis interino (un número de al menos 200 eventos se asoció a menor riesgo de sesgo en este sentido)<sup>(2)</sup>.

Muchos autores recomiendan tomar la decisión de la interrupción de hallarse beneficio basados en un nivel menor de valor de p (ejemplo ajustar el nivel de p de acuerdo a la cantidad de análisis interinos pre-especificados, entonces si se planean tres análisis interinos el valor de p para asumir que existen diferencias entre las intervenciones sería 0,0221)<sup>(6)</sup>. Sin embargo hay que tener en cuenta que disminuir el valor de p no excluye en forma segura la posibilidad de hallazgos por error si el número de eventos es pequeño.

También será relevante valorar resultados de otros estudios que intentan responder la misma pregunta para poner en contexto los hallazgos de un estudio que se interrumpió antes de la finalización del mismo.

Un ejemplo claro del riesgo que se asume ante resultados de análisis interinos muy beneficiosos es analizado por Wheatley y cols<sup>(7)</sup>, que describen los análisis interinos de un ECCA que reclutó pacientes con leucemia mieloide aguda aleatorizados a dos estrategias de quimioterapia: 4 versus 5 ciclos para la reducción de mortalidad<sup>(8)</sup>. Los investigadores observaron tras el primer análisis interino (en el año 1998 luego de haber reclutado 340 pacientes y observado 65 muertes) un claro beneficio del curso de 5 ciclos sobre el comparador (OR 0,47; P=0,0003). Sin embargo, a pesar de que este resultado interino sobrepasaba el límite pre-especificado para la interrupción temprana del estudio, los investigadores decidieron continuar con el desarrollo del mismo, ya que consideraron que el efecto observado era muy superior al que podían esperar. Luego de finalizado el reclutamiento del estudio (1078 pacientes y 297 muertes), los investigadores observaron que el OR se modificó invirtiendo el sentido del efecto inicial (OR=1,09) determinando que no hay diferencias entre las dos estrategias de quimioterapia e inclusive pudiendo ser perjudicial la administración de 5 ciclos sobre 4.

### Conclusiones y consideraciones ante estudios interrumpidos antes de tiempo por beneficio

Cuando estemos frente a ECCAs interrumpidos antes de tiempo por beneficio debemos ser cautelosos con la interpretación de los resultados. En primera instancia deberíamos rastrear la literatura en busca de otros estudios que respondan la misma pregunta y, si es posible, revisiones sistemáticas que valoren los efectos de la intervención en múltiples estudios. Deberemos ser cuidadosos en leer la sección de Métodos buscando bajo qué condiciones se decidió la interrupción del ECCA (límites pre-especificados, justificación, comité independiente de análisis de datos). Si esto no se describe adecuadamente debemos disminuir nuestra grado de confianza en la validez de los resultados.

Aún si el plan de análisis estadístico y el plan de acción estuvieran bien pre-especificados y detallados debemos recordar que los estudios que se interrumpen prematuramente por beneficio tienen un riesgo elevado de sobrestimar el efecto beneficioso en al menos un 30%<sup>(9)</sup>.

**Bibliografía**

1. Bottaro FJ. ¿Cómo leer la literatura científica? Analizando la evidencia que evalúa intervenciones terapéuticas. (Parte V). *Hematología*. 2014;18(3): 293-298.
2. Montori VM, Devereaux PJ, Adhikari NK y col. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294(17):2203-9.
3. Murad MH, Guyatt GH, Domecq JP y col. Randomized trials addressing a similar question are commonly published after a trial stopped early for benefit. *J Clin Epidemiol*. 2016 Nov 7. pii: S0895-4356(16)30585-6. doi:10.1016/j.jclinepi.2016.10.006. [Epub ahead of print]
4. Bassler, D., Montori, V., Devereaux, P y cols. (2015) Randomized trials stopped early for benefit. In: Guyatt G, Drummond R, Meade MO, Cook DJ, editors. *User's Guide To The Medical Literature. A Manual for Evidence-Based Clinical Practice*. JAMA & Archives Journals, American Medical Association; 2015. (3<sup>rd</sup> edition, pp. 134-141).
5. Guyatt G, Jaeschke R, Heddle N y col. Basic statistics for clinicians 2, Interpreting study results: confidence intervals. *CMA*. 1995;152:169-73.
6. Pocock SJ. When to stop a clinical trial. *BMJ*. 1992;305:235-240.
7. Wheatley K, Clayton D. Be skeptical about unexpected large apparent treatment effects: the case of an MRC AML12 randomization. *Control Clin Trials*. 2003;24(1):66-70.
8. Wheatley K, Burnett AK, Gibson B y col. Optimising consolidation therapy: four versus five courses, SCT versus chemotherapy—preliminary results of MRC AML12. *Hematology J*. 2002;3(Suppl 1):159.
9. Bassler D, Briel M, Montori VM y col. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA*. 2010;303(12):1180-7.