

## Análisis crítico de ensayos clínicos. ¿Es oro todo lo que reluce?

A. Etxeberria Aguirre<sup>a\*</sup>, R. Rotaecche del Campo<sup>b\*</sup> y J. M. Morán Barrios<sup>c\*</sup>

<sup>a</sup> Servicio de Farmacia de Atención Primaria. Comarca Gipuzkoa Este-Ekialde. San Sebastián. Osakidetza-Servicio Vasco de Salud. <sup>b</sup> Centro de Salud de Alza. San Sebastián. Guipuzcoa. España. <sup>c</sup> Unidad de Investigación y Docencia. Hospital de Cruces. Baracaldo. Vizcaya. Osakidetza-Servicio Vasco de Salud. España

El volumen actual de la literatura médica es inabarcable y no toda la información disponible es relevante y válida. Un caso particular es el del ensayo clínico aleatorizado (ECA) que se considera el patrón de oro para medir la eficacia de una intervención (farmacológica o no). El enfoque de la medicina basada en la evidencia insiste en la importancia de la evaluación crítica de los ECA en cuanto a validez, impacto clínico y aplicabilidad. El profesional sanitario precisa utilizar la lectura crítica de la literatura médica para ser autónomo en sus decisiones. Para la interpretación de los ECA se requiere conocer algunas claves de su diseño, como son la correcta asignación aleatoria, el enmascaramiento o el cálculo del tamaño muestral, así como reflexionar sobre otros aspectos más sutiles (pero más intuitivos para el clínico) relacionados con la relevancia de la pregunta de investigación y con la forma de presentación de los resultados. En este artículo se revisan algunos conceptos como el sesgo de publicación, el sesgo en la presentación de resultados, la elección de variables (agregadas, principales o secundarias, subrogadas o finales) y cómo se miden los resultados. Finalmente debemos tener presente que el proceso de razonamiento clínico y de toma de decisiones es mucho más complejo que la mera aplicación de los resultados de la investigación, que siempre han de integrarse con la experiencia clínica individual, las características y preferencias de cada paciente y el contexto sanitario.

Palabras clave: ensayos clínicos, sesgos, lectura crítica, medicina basada en la evidencia.

### Critical analysis of clinical trials. Is everything that shines gold?

It is impossible to cover the current volume of medical literature and not all the information available is relevant and valid. A specific case is that of the randomized clinical trial (RCT) that is considered the gold standard to measure efficacy of an intervention (pharmacological or not). Evidence based medicine approach stresses the importance of critical evaluation of the RCTs in regards to validity, clinical impact and applicability. The health care professional must make a critical reading of the medical literature to be independent in his/her decisions. In order to interpret the RCTs, some key factors of their design, such as correct random allocation masking or calculation of the sample size, and reflecting on other more subtle aspects (but more intuitive for the clinician), related with the relevance of the research question and presentation from of the results must be known. In this review, some concepts such as publication bias, result presentation bias, choice of endpoints (aggregate, primary or secondary, subrogated or final) and how the results are measured are reviewed. Finally, we should keep in mind that the clinical reasoning process and decision making one is much more complex than the mere application of the research results that must always be accompanied by individual clinical experience, the characteristics and preferences of each patient and health care context.

Key words: clinical trial, biases, critical reading, evidence based medicine.

### Introducción

Los profesionales sanitarios debemos resolver múltiples cuestiones para proporcionar los mejores

cuidados a nuestros pacientes. Entre ellas, las preguntas sobre la elección del mejor tratamiento ocupan un lugar destacado<sup>1</sup>. El tipo de publicación más adecuado para contestar las preguntas sobre la eficacia de una intervención (ya sea un fármaco antihipertensivo, una estrategia preventiva, un programa educativo o un método de cribado) es el ensayo clínico aleatorizado (ECA)<sup>2</sup>. A su vez, la realización de ensayos clínicos es un requisito legal para poder registrar y comercializar un nuevo fármaco o autorizar una nueva indicación, por lo que la mayoría de los ECA con fármacos están promovidos por compañías farmacéuticas.

#### Correspondencia:

J. M. Morán Barrios.  
Unidad de Investigación y Docencia.  
Hospital de Cruces.  
48903 Baracaldo. Vizcaya. España.  
Correo electrónico: jesusmanuel.moranbarrios@osakidetza.net

Recibido: 5 de febrero de 2007.

Aceptado: 18 de junio de 2007.

\*Miembros del grupo CASPe del País Vasco.

Cada año Medline cataloga 560.000 artículos nuevos y Cochrane Central añade aproximadamente 20.000 ensayos aleatorizados nuevos, lo que supone la publicación de 55 nuevos ECA cada día. Estar al día se ha convertido en un verdadero reto para el profesional sanitario. El campo de las enfermedades cardiovasculares y el de la hipertensión (HTA) en particular destacan especialmente en la publicación de ECA. Una búsqueda de ECA con los términos "hipertensión" en Medline durante el año 2006 nos ofrece más de 600 referencias. Además, hoy en día accedemos desde nuestro ordenador a los artículos originales publicados en múltiples revistas. Este escenario nos exige el desarrollo de nuevas habilidades para "estar al día" como puede ser la búsqueda bibliográfica.

Aunque seamos capaces de identificar la información que necesitamos, tenemos que saber interpretarla adecuadamente. Este proceso se conoce como "evaluación crítica" o "lectura crítica" de la literatura médica. Esta labor, junto con nuestra experiencia personal y las preferencias de nuestros pacientes, nos ayudará a tomar las decisiones apropiadas en un contexto clínico concreto. Sólo recientemente la evaluación crítica de la literatura médica se ha introducido en los currículos profesionales, siendo el movimiento de la medicina basada en la evidencia (MBE) el que ha subrayado la importancia de la evaluación crítica de los ECA en lo que se refiere a la validez, impacto clínico y aplicabilidad<sup>3</sup>. De hecho sabemos que no toda la información disponible es relevante, válida y fiable<sup>4</sup>.

Hoy día supone un verdadero reto el que los clínicos puedan incorporar los resultados de los ECA relevantes y de calidad al proceso de razonamiento clínico y de toma de decisiones en un contexto de una fuerte carga asistencial y de una influyente presión comercial que rodea a este tipo de estudios. ¡Y todo ello sin necesidad de ser un experto en estadística!

A lo largo del artículo iremos viendo que a los ya "clásicos" sesgos relacionados con la propia metodología del ensayo clínico, como la distribución aleatoria de los grupos, el enmascaramiento o el cálculo del tamaño de la muestra, se añaden otras dificultades como las relacionadas con la relevancia de la pregunta de investigación y en la forma en que se presentan los resultados<sup>5</sup>. La experiencia clínica y la visión de los pacientes son cruciales para decidir si la pregunta que se intenta responder con un ensayo clínico es pertinente o si las medidas de resultado que utiliza un estudio para valorar la eficacia de un nuevo antihipertensivo son realmente las que elegirían nuestros pacientes.

El objetivo de este artículo es proporcionar a los clínicos algunas claves para poder ser más autónomos a la hora de examinar un ECA e interpretar sus resultados.

## ¿Qué es y para qué sirve un ensayo clínico aleatorizado?

Un ECA es un estudio experimental en el que se aplica una intervención (por ejemplo, un fármaco para reducir las cifras de presión arterial o una dieta sin sal) a un grupo de pacientes (grupo experimental) y el resultado (por ejemplo, los niveles de presión arterial o la frecuencia de eventos isquémicos cardiovasculares) se compara con el obtenido en un grupo idéntico de pacientes (grupo control) que recibe otra intervención (por ejemplo, otro fármaco, un placebo o la dieta habitual). Los pacientes son asignados de forma aleatoria al grupo experimental y al grupo control con la misma probabilidad y al mismo tiempo de forma que ambos grupos sean similares en todo (características de los pacientes, factores pronósticos conocidos o no, otros tratamientos recibidos, etc.), es decir, grupos homogéneos excepto en la intervención que reciben. De esta forma, la diferencia observada podrá atribuirse al tratamiento recibido, con un margen de error establecido, y siempre que el ensayo esté libre de sesgos<sup>2</sup>.

## ¿Qué aspectos determinan el "qué" y el "cómo" se publica un ensayo clínico? Sesgo de publicación y sesgo en la presentación de los resultados

No todos los ensayos clínicos se publican. Los investigadores y los financiadores son menos proclives a enviar a las revistas los estudios con resultados estadísticamente no significativos (estudios "negativos") y los editores tradicionalmente han preferido publicar los ensayos "positivos"<sup>2</sup>. Esta tendencia puede provocar una percepción sesgada de la eficacia de una intervención. En un estudio<sup>6</sup> los autores hicieron un seguimiento de los 42 ensayos clínicos de 5 diferentes antidepresivos inhibidores selectivos de la recaptación de serotonina (ISRS) que habían sido presentados a la Agencia Sueca del Medicamento entre 1983-1994. De los 21 ensayos en que el fármaco fue más eficaz que el placebo, 19 fueron publicados como ECA, mientras que de los otros 21 ensayos que no mostraron diferencias, sólo 6 fueron publicados. El sesgo de publicación también puede darse publicando los trabajos varias veces (en varias revistas, separatas, realizando distintos metaanálisis con los mismos estudios, etc.), provocando la sensación de que existe más evidencia de la que realmente hay<sup>4,5,7</sup>.

Por otro lado, las revisiones sistemáticas han demostrado que los ECA financiados por compañías farmacéuticas tienen entre 4 y 5 veces más probabilidades de obtener resultados favorables

a la intervención que los ensayos con otras fuentes de financiación (organismos públicos o entidades sin ánimo de lucro)<sup>8,9</sup>, siendo excepcionales los ensayos financiados por la industria que favorecen la terapia tradicional frente a sus productos<sup>7</sup>. Esta clara tendencia a favorecer al producto investigado por la industria no se debe a una peor calidad metodológica<sup>8</sup>. Los motivos están más relacionados con el propio objetivo del estudio (por ejemplo, comparando el producto experimental frente a placebo en lugar del fármaco de referencia) o midiendo múltiples resultados y publicando sólo los favorables<sup>8,9</sup>.

Las iniciativas recientes de crear un registro de libre acceso para los ECA a su inicio (<http://clinicaltrials.com>) o la decisión del Comité Internacional de Editores de Revistas Médicas de publicar sólo aquellos ensayos clínicos que han sido previamente registrados pueden contribuir a disminuir el sesgo de publicación. La nueva Ley 29/2006 de garantías y uso racional de los medicamentos contempla también la creación de un registro nacional público y de libre acceso, así como la obligatoriedad de que los promotores publiquen los resultados del ensayo clínico, sean positivos o no.

### **¿Es el ensayo relevante y aplicable para mis pacientes? La importancia de reconsiderar la pregunta de investigación**

Antes de leer un artículo merece la pena valorar si el objetivo que plantea el ensayo es relevante y aplicable a nuestros pacientes<sup>10</sup>. Ello implica preguntarse hasta qué punto los pacientes a estudio y las intervenciones que se van a comparar se parecen a nuestros pacientes y a las terapias que disponemos y si los resultados o outcomes que mide el estudio son importantes para nuestros pacientes<sup>11</sup>. La relevancia implica plantearnos hasta qué punto la pregunta que plantea el estudio podría modificar nuestra práctica clínica. La validez externa o la aplicabilidad se refieren a en qué medida los resultados del estudio pueden ser extrapolables a una población más amplia o a unos pacientes en particular<sup>3</sup>. De hecho, no más del 4%-15%<sup>12</sup> de los pacientes valorados para su inclusión son finalmente aleatorizados en una mayoría de los ECA<sup>3</sup>.

Habitualmente los ECA miden la eficacia de las intervenciones en condiciones "ideales" con pacientes muy homogéneos, criterios de inclusión y exclusión muy estrictos (tienden a excluir a niños, pacientes muy ancianos, con comorbilidad importante o polimedicados) y con un estrecho seguimiento de los pacientes distinto a las condiciones de la práctica clínica habitual. En el estudio RALES<sup>13</sup> la espironolactona redujo

la mortalidad y las hospitalizaciones en pacientes con insuficiencia cardíaca severa y función renal normal. Posteriormente, en estudios observacionales el uso de este fármaco en pacientes con disfunción renal severa y/o insuficiencia cardíaca leve se ha asociado a un incremento de muertes y hospitalizaciones por hiperpotasemia<sup>14</sup>.

### **¿Es adecuada la elección de comparador?**

Los posibles beneficios y riesgos de las terapias experimentales deberían ser evaluados mediante su comparación con las mejores terapias estándar recomendadas. El uso del placebo como comparador es una práctica tan frecuente como polémica. Otras formas para diseñar estudios que favorezcan al producto experimental consisten en comparar el fármaco frente a uno que ya de antemano se prevea inferior, como algunos autores<sup>15</sup> han sugerido a propósito del ensayo ASCOT<sup>16</sup>; frente a una dosis baja del competidor (como atorvastatina, 80 mg, frente a pravastatina, 40 mg) o frente a una dosis alta del competidor cuando se trata de demostrar un mejor perfil de efectos adversos<sup>7</sup>.

### **La evaluación de los resultados: ¿se basa en las variables importantes para los pacientes? Variable final frente a intermedia o "subrogada", variables principales, secundarias y agregadas**

#### **¿Variables subrogadas o finales?**

Una variable intermedia o "subrogada" es un resultado fisiológico, bioquímico o de otro tipo, relativamente sencillo de obtener (como las cifras de presión arterial), que se usa cuando los eventos clínicos importantes (infartos, mortalidad, etc.), llamados variables finales, son difíciles de observar, ya que implican estudios de larga duración con un gran número de pacientes<sup>3,4</sup>. El valor de una variable subrogada dependerá de la medida en que sea buena predictora de la variable final. Por ejemplo, la terapia hormonal sustitutiva redujo los niveles de colesterol LDL<sup>17</sup> (variable intermedia), pero posteriormente se demostró que aumentaba el riesgo de eventos cardiovasculares<sup>18</sup> (variable final).

#### **¿Variable principal o secundaria?**

Los ensayos clínicos están diseñados para responder a una pregunta concreta, por ejemplo, si un fármaco es más eficaz que otro en prevenir un evento específico, que se denomina "variable



ble de resultado principal<sup>14</sup>. La variable principal nos indica el objetivo principal del estudio y, por tanto, el tamaño de la muestra se calcula en función de ésta. Sería raro que una sola variable abarcara todos los aspectos que interesan a los investigadores, por lo que habitualmente los ensayos miden también otros resultados “secundarios” con fines más descriptivos para formular hipótesis o que pueden ser incluidos en un metaanálisis junto con los resultados de otros ensayos. Las variables secundarias deben estar contempladas en el diseño del estudio y adecuadamente justificadas. Además es posible que el estudio no tenga una potencia estadística adecuada para las variables secundarias<sup>4</sup>, por lo que la ausencia de una diferencia estadísticamente significativa no debe interpretarse como evidencia de que tales diferencias no existen. Por otro lado debemos ser muy cautos ante los resultados inesperados o discordantes procedentes de variables secundarias, particularmente en los análisis *post hoc* o cuando el ensayo no ha mostrado diferencias en la variable principal. En el ensayo ELITE I<sup>19</sup>, que incluía a 722 ancianos con insuficiencia cardíaca, no se observaron diferencias entre captopril y losartán en la función renal (variable principal). Sin embargo, se detectó una disminución inesperada en la mortalidad a favor de losartán. Este hallazgo dio lugar al ensayo ELITE II<sup>20</sup>, con un diseño prácticamente idéntico pero incluyendo a 3.152 pacientes y con la mortalidad como variable principal. No se observaron diferencias entre losartán y captopril.

**Variables agregadas (“composite endpoint”)**

El uso de variables agregadas o compuestas (que engloban distintos efectos) es cada vez más frecuente, especialmente en el área cardiovascular. Para el investigador tienen la ventaja de que requieren menor tamaño de muestra y menor tiempo de seguimiento; para el clínico la desventaja de que son más propensas a ser malinterpretadas<sup>3</sup>. La validez de una variable agregada depende de que la intervención afecte de igual forma a todos los componentes, la frecuencia con la que ocurren y la importancia clínica que tengan para los pacientes<sup>21</sup>. Imaginemos que debemos aconsejar el tratamiento más adecuado para un paciente anciano con angina sintomática no controlada farmacológicamente en base a los hallazgos del estudio TIME<sup>22</sup>. Este ECA comparó el tratamiento quirúrgico frente a la terapia médica en pacientes mayores de 75 años con angina sintomática no controlada con dos fármacos, utilizando como medida de eficacia la variable agregada de muerte, infarto no fatal y hospitalización por síndrome coronario agudo (tabla 1). En una primera lectura el tratamiento quirúrgico parece

**TABLA 1**  
**Resultados del ensayo TIME. Tratamiento invasivo frente a médico en la angina no controlada en población mayor de 75 años**

| RESULTADO                                    | INVASIVO<br>(n = 153) | MÉDICO<br>(n = 148) | DIFERENCIA<br>ABSOLUTA<br>DE RIESGO<br>% (IC 95%) | HR<br>(IC 95%)        |
|--|-----------------------|---------------------|---|-----------------------|
| Variable agregada                            | 39<br>(25,5%)         | 95<br>(64,2%)       | 38,7%<br>(27,9-48,5)                              | 0,31<br>(0,21 a 0,45) |
| Muerte                                       | 17                    | 12                  | -3,0%<br>(-9,9 a 3,8%)                            | 1,51<br>(0,72 a 3,16) |
| Infarto no fatal                             | 14                    | 20                  |   | 0,75<br>(0,36 a 1,66) |
| Hospitalización por síndrome coronario agudo | 28                    | 106                 |   | 0,19<br>(0,12 a 0,30) |

HR: *hazard ratio*; IC: intervalo de confianza.

ser mucho más eficaz. Sin embargo, si consultamos detenidamente los criterios sobre la interpretación de variables agregadas<sup>21</sup> (tabla 2) es posible que nuestra actitud sea más conservadora que la que nos podía sugerir un vistazo rápido a las conclusiones del estudio.

**¿Se valora la seguridad de las intervenciones de forma adecuada y durante el tiempo necesario?**

A diferencia de los resultados sobre eficacia, en los ECA con fármacos las variables de seguridad

**TABLA 2**  
**Guía para interpretar las variables agregadas**

1. ¿Los distintos componentes tienen la misma importancia clínica? La muerte y el infarto tienen mayor importancia clínica que las hospitalizaciones por síndrome coronario
2. ¿Ocurren los resultados con similar frecuencia? Las hospitalizaciones ocurren con mucha mayor frecuencia que la muerte o los infartos
3. ¿El efecto del tratamiento sobre los distintos componentes es similar? (valorar la reducción del riesgo y los intervalos de confianza y el sustrato biológico y el razonamiento clínico) La reducción del riesgo ocurre sobre todo en las hospitalizaciones, el efecto sobre el infarto no es significativo y se observa una tendencia a aumentar la mortalidad (*a priori* era esperable que el tratamiento quirúrgico aumentara la mortalidad perioperatoria). Los intervalos de confianza son muy amplios para la muerte e infarto

raramente constituyen el objetivo principal del estudio. Los pacientes con mayor riesgo de sufrir efectos adversos tienden a ser excluidos de los ensayos. Además, a menudo los datos sobre efectos adversos se describen de forma inadecuada en las publicaciones<sup>23</sup>. La consecuencia inmediata es que cuando se comercializa un nuevo fármaco la información acerca de su seguridad suele ser incompleta y el balance riesgo-beneficio desconocido. Basta recordar los casos de rabdomiólisis por cerivastatina, los infartos por rofecoxib o los efectos perjudiciales de la terapia hormonal sustitutiva.

Se recomienda, por tanto, una buena dosis de cautela a la hora de aceptar los resultados de los ECA que no describen adecuadamente los efectos adversos o que no están adecuadamente diseñados para detectarlos, ya que nos pueden dar una falsa sensación de seguridad.

La duración de un ensayo es otro aspecto clave a la hora de interpretar sus resultados, y conviene preguntarse siempre ¿qué ocurriría si el ensayo durara más tiempo? En el reciente ensayo CIDS<sup>24</sup>, que compara la implantación de un desfibrilador con el tratamiento con amiodarona durante 5 meses para la prevención de muerte súbita, no se observaron diferencias entre ambas opciones. Sin embargo, uno de los centros continuó el ensayo con 120 pacientes durante 5,6 años, observándose una mayor mortalidad en el grupo tratado con amiodarona, posiblemente atribuible a las arritmias producidas por el fármaco<sup>25</sup>.

### **¿Es el ensayo metodológicamente correcto? Tres cuestiones para decidir si merece la pena seguir usando nuestro tiempo: la distribución aleatoria, el enmascaramiento y el análisis por intención de tratar**

Los aspectos más "técnicos" de los ECA, como la forma en que se realiza la distribución aleatoria de los grupos o el seguimiento de los pacientes, son probablemente los más difundidos en las publicaciones sobre diseño de ECA. Existe evidencia empírica de que estos aspectos pueden sobreestimar de forma importante la eficacia de una intervención cuando el método seguido no es el adecuado<sup>26</sup>. La declaración CONSORT<sup>27</sup> (*Consolidated Standards of Reporting Trials*) ha supuesto un avance importante para que la publicación de los ECA sea más completa y recoja de forma adecuada los aspectos metodológicos.

#### **Aleatorización**

La fortaleza principal de un ensayo clínico es la adecuada aleatorización. Es importante que ha-

ya ocultación de la secuencia de aleatorización (OSA), es decir, que el investigador no puede saber qué rama le va a corresponder a cada sujeto ni pueda influir de manera consciente o inconsciente en el orden de aleatorización. El estudio CAPPP es un buen ejemplo de las implicaciones de la ausencia de OSA en el diseño de un ECA<sup>28</sup>. Este estudio comparó la frecuencia de eventos cardiovasculares de captopril frente a la terapia convencional con tiazidas y/o metoprolol en una muestra de 11.000 hipertensos seguidos durante 6 años. La aleatorización en dos grupos se hizo mediante sobres cerrados. Con ese número de pacientes lo esperable era que ambos grupos fueran idénticos en cuanto a las variables pronósticas iniciales. Sin embargo, ambos grupos diferían sustancialmente: los pacientes aleatorizados a captopril tenían una presión arterial sistólica (PAS) 2,2 mmHg más elevada y una frecuencia mayor de diabetes. Estos hallazgos sugieren que algunos de los médicos implicados en el proyecto conocían el contenido de los sobres (posiblemente no serían opacos) y seleccionaron con más frecuencia captopril para pacientes diabéticos o con cifras mayores de presión arterial (PA). La consecuencia de ello es que los resultados del ensayo son difícilmente interpretables.

En los estudios multicéntricos la OSA se puede garantizar, por ejemplo, generando la secuencia aleatoria por ordenador mediante la utilización de un sistema interactivo de respuesta por voz (IVRS). La ausencia de OSA puede sobreestimar el efecto de las intervenciones hasta en un 40%<sup>26</sup>.

#### **Enmascaramiento**

Idealmente, en un ECA ni el paciente ni el investigador deberían saber qué tratamiento está recibiendo el grupo experimental o el grupo control, lo que se conoce como doble ciego. Técnicamente implica que las dos opciones a comparar sean idénticas en apariencia y se administren de la misma forma. El enmascaramiento es especialmente importante cuando las variables que se miden son subjetivas (como calidad de vida, dolor, etc.) o cuando es posible que, al margen de la intervención a estudio, puedan añadirse otros tratamientos concomitantes (el investigador podría añadirlos de forma desigual al grupo experimental y al control). El enmascaramiento protege del sesgo de observación y del sesgo producido por tratar de forma diferente a ambos grupos, al margen de las intervenciones a estudio. Si por razones éticas o técnicas el ensayo no puede ser doble ciego (por ejemplo, intervenciones quirúrgicas, educativas, tratamientos parenterales, programa de ejercicio físico, pautas de dosificación muy complejas, etc.), al menos el evaluador de los resultados debería

desconocer la intervención asignada (evaluación ciega por terceros) y utilizar medidas de resultado objetivas (como los resultados de laboratorio, mortalidad, infarto, etc.). Nos deberemos fijar en si, al margen de la intervención, los grupos han sido tratados de la misma forma.

### El seguimiento de los pacientes

Siempre se deben cuantificar y describir las causas de la pérdida de pacientes. Es importante que los motivos de las pérdidas afecten de la misma forma al grupo experimental y al control. Si en una de las ramas hay más abandonos debido a efectos adversos o a falta de eficacia, los resultados estarán sesgados; si las causas de las pérdidas se distribuyen de la misma forma en las dos ramas, su influencia será menor. Es importante que al analizar los resultados se siga el principio del análisis por intención de tratar (ITT), de forma que todos los pacientes se contabilicen en la rama a la que fueron aleatorizados, aunque no completen el estudio. Se considera que el ITT produce unos resultados más parecidos a los de la realidad, es decir, los resultados son más generalizables, preserva la aleatorización al mantener el equilibrio entre los grupos respecto a los factores pronósticos, disminuye la influencia de los abandonos del tratamiento y la de los pacientes no cumplidores, y además es más conservador<sup>29</sup>. La excepción a esta regla son los estudios de equivalencia y de no inferioridad donde se deben presentar tanto el ITT como el análisis por protocolo<sup>30</sup>.

### ¿Cómo es la magnitud y precisión de los resultados? La *p* frente a los intervalos de confianza, los riesgos relativos frente a los absolutos

#### Potencia estadística, *p* e intervalo de confianza

En términos sencillos la *p* (nivel de significación estadística) sólo nos mide la probabilidad de que los resultados sean explicados por el azar si en realidad no hubiera diferencias entre los tratamientos. De manera arbitraria se considera que una  $p < 0,05$  es lo suficientemente pequeña como para aceptar que las diferencias se deben a la intervención y no al azar. Tradicionalmente se le ha dado mucha importancia, pero el valor de *p* no dice nada acerca de la magnitud del efecto. Podemos tener una *p* muy significativa, pero una diferencia clínicamente irrelevante<sup>4</sup>. La potencia estadística es la probabilidad de detectar una diferencia significativa real. Arbitrariamente se considera adecuado un poder estadístico del 80% para la mayoría de los estudios.

Antes de realizar un ensayo se debe realizar el cálculo del tamaño de muestra teniendo en cuenta el riesgo alfa, la potencia, la magnitud del efecto que se considera clínicamente importante, la frecuencia esperada de eventos en el grupo control. Si el número de pacientes es demasiado pequeño es muy posible que no encontremos diferencias significativas, aunque éstas realmente existan. Al contrario, aumentando desproporcionadamente el tamaño de la muestra casi siempre será posible demostrar una diferencia estadísticamente significativa, aunque clínicamente sea irrelevante.

El intervalo de confianza (IC) indica el rango de valores entre los que se encontrará el verdadero valor de la población, con un nivel de confianza que se establece de antemano (generalmente el 95%). El IC nos indica la magnitud del efecto y también la precisión de las estimaciones<sup>4</sup>. En el estudio TIME<sup>22</sup> (tabla 1) el tratamiento quirúrgico presentaba un riesgo tres veces menor que el médico (HR 0,31), con un intervalo relativamente estrecho (preciso). Sin embargo, el IC para la mortalidad puede ir desde 0,72, compatible con un efecto beneficioso, hasta 3,16, compatible con un riesgo de muerte tres veces mayor.

#### Riesgo relativo frente a riesgo absoluto, ¿qué nos dicen los números?

Imaginemos que tenemos a un paciente con un riesgo cardiovascular elevado, candidato a recibir una intervención farmacológica que disminuya su riesgo. Consideremos tres posibles escenarios:

- 1) Tras 9 meses de tratamiento, los pacientes que reciben el fármaco A tienen un riesgo un 20% menor de padecer un evento cardiovascular que los pacientes que reciben el fármaco B.
- 2) Tras 9 meses de tratamiento el 9,2% de los pacientes que reciben el fármaco A presentan un evento cardiovascular frente al 11,4% de los pacientes que reciben el fármaco B.
- 3) Por cada 48 pacientes tratados con el fármaco A en lugar del fármaco B durante 9 meses evitaremos un evento cardiovascular, pero por cada 100 pacientes tratados con A en lugar de B durante el mismo período uno sufrirá un sangrado adicional.

¿Cuál de las tres opciones preferimos para nuestro paciente? Muchos lectores se sorprenderían al conocer que los tres escenarios representan datos del mismo ensayo y que los números no son más que una "ilusión", distintas formas de presentar los mismos resultados. El ensayo en cuestión es el CURE<sup>31</sup>, que comparó la combinación de clopidogrel + aspirina (A) frente a aspirina sola (B) en el síndrome coronario agudo. Los resultados de los ECA se presentan más a menudo como riesgo relativo (RR) y reducción



relativa del riesgo (RRR). El RR estima el efecto de una intervención (tasa de eventos del grupo de intervención) dividida por el efecto de su comparador (tasa de eventos del grupo control) ( $9,3/11,4 = 0,82$ ) y la RRR un 18% (1-RR). En el estudio CURE el riesgo absoluto de padecer el resultado principal en el grupo tratado con la asociación fue del 9,3% y del 11,4% en el grupo de la aspirina. La reducción absoluta de riesgo (RAR) es la diferencia de la frecuencia de eventos entre los dos grupos ( $11,4\% - 9,3\% = 2,1\%$ ), y ésta dividida entre la tasa de eventos del grupo control ( $2,1/11,4 = 0,18$  ó 18%) nos indica también la RRR.

El NNT (*number needed to treat*) es el número de pacientes que se necesitaría tratar con una intervención específica para producir o evitar un evento adicional. Se calcula como  $1/RAR$  (la RAR la expresaremos entre 0 y 1, no en porcentaje). Razonado como regla de tres, si tratando a 100 pacientes evitamos 2,1 eventos principales, para evitar 1 evento trataremos a 48. Cuanto menor sea el NNT mayor será el efecto de la intervención. El NNT depende del riesgo basal, ya que las intervenciones dirigidas a personas de alto riesgo tienen mayor beneficio absoluto que las que se dirigen a poblaciones de bajo riesgo. El NNT se debe referir siempre a un tiempo determinado y a una población determinada. Por ejemplo, si el riesgo de muerte de la enfermedad A es del 7% (tasa de muerte en el grupo control) y con el tratamiento experimental es del 2%, la RAR sería del 5% y el NNT de 20 ( $1/0,05$ ). Por el contrario, si el riesgo de muerte de la enfermedad B es de 0,7% y con el tratamiento (grupo intervención) es del 0,2%, la RAR sería del 0,5% y el NNT de 2. Pero para ambos estudios el RR y la RRR son iguales (29% y 71%, respectivamente), por tanto la toma de decisiones debería ir en función de la RAR y del NNT.

El NNH (número necesario para perjudicar o *number needed to harm*) se define del mismo modo que el NNT, pero para medir efectos indeseables. Para el sangrado el NNH sería ( $1/[0,037 - 0,027]$ ) = 100. En nuestro ejemplo podemos concluir que por cada 100 pacientes tratados con aspirina + clopidogrel en lugar de aspirina sola en el síndrome coronario agudo durante 9 meses evitaremos 2 eventos (muerte, infarto o ictus) y produciremos una hemorragia. Visto así, el balance entre beneficios y riesgos es aún favorable, aunque seguramente menos espectacular que pensar que evitaremos el 20% de los infartos, ictus o muertes. De hecho, existe evidencia de que son menos los médicos y pacientes que eligen el tratamiento experimental si los datos se presentan como reducción absoluta del riesgo y NNT que si son presentados como RR o RRR<sup>32,33</sup>.

TABLA 3

**Criterios de validez en el análisis de subgrupos**

|  |
|--|
| La hipótesis específica se había planteado en el diseño del estudio y es una de las pocas hipótesis que se quiere contrastar |
| La magnitud del efecto es grande   |
| El efecto del tratamiento es altamente significativo (cuanto más bajo es el valor de la $p$ más creíble será la diferencia)  |
| Los resultados son consistentes con los de otros estudios  |
| Existe evidencia indirecta a favor de los resultados encontrados (plausibilidad biológica)                                   |
| Más creíbles si se observa un beneficio en la población global del estudio   |

### ¿Qué son los análisis de subgrupos? La tentación de la manipulación de los resultados. Los análisis a posteriori

El análisis de subgrupos tiene como objetivo identificar grupos de pacientes en los cuales el beneficio de la terapia experimental es mayor o menor que la del promedio de todos los pacientes del ensayo. Los subgrupos pueden definirse en función de la edad, sexo, gravedad, comorbilidad u otras intervenciones concomitantes<sup>3</sup>. Aunque este tipo de análisis pueda parecer interesante, presenta serios problemas de multiplicidad, de forma que si realizamos un número "suficiente" de análisis es muy probable obtener algún resultado significativo (falso positivo) debido al azar<sup>34</sup>. La revista *Lancet* publicó un ejemplo muy ilustrativo de los sesgos en la interpretación del análisis de subgrupos. En el ensayo ISIS-2, tras realizar 40 análisis de subgrupos, quedó "demostrado" que para los pacientes nacidos bajo los signos Géminis y Libra del zodiaco el tratamiento con aspirina era perjudicial<sup>34</sup>.

En la tabla 3 se presentan unos criterios a tener en cuenta antes de basar las decisiones en un análisis de subgrupos<sup>35</sup>. Diversos autores desaconsejan el análisis de subgrupos. En su lugar cuando sea necesario valorar el efecto de las características basales de los pacientes sobre la eficacia de una intervención es preferible realizar test estadísticos de interacciones<sup>34</sup>.

### Consideraciones finales. ¿Cómo podemos ser autónomos en la lectura de ensayos clínicos aleatorizados?

El clínico interesado en proporcionar los cuidados más adecuados a sus pacientes debe combinar su experiencia con los resultados proporcionados por la mejor investigación clínica. Es

**TABLA 4**  
**Guía para evitar la interpretación sesgada de los datos**

|   |
|---|
| <p>Leer únicamente las secciones de "método" y "resultados", obviando la sección de discusión</p> <p>Fijarse en los comparadores inadecuados: ¿es el placebo el comparador adecuado?</p> <p>Fijarse en la elección de las variables de resultado: ¿agregadas?, ¿subrogadas?, ¿secundarias?</p> <p>Cuidado cuando el efecto del tratamiento (la magnitud) es pequeño (por ejemplo, estudios con gran número de pacientes con diferencias significativas próximas al efecto nulo). Valorar los resultados en términos absolutos (NNT, reducción absoluta del riesgo) más que en medidas relativas como el riesgo relativo o la reducción relativa del riesgo. Fijarse en los límites de los intervalos de confianza</p> <p>Leer los resúmenes evaluados críticamente en revistas secundarias (como <i>ACP Journal Club</i> o <i>Evidence-based Medicine</i>) realizados por un equipo de clínicos y metodólogos sin conflictos de interés</p> <p>Atención a los análisis de subgrupos</p> |
|---|

NNT: *number needed to treat*. Modificada de Montori VM<sup>36</sup>.

habitual que los ECA más novedosos se difundan profusamente a través de simposios, congresos o separatas de revistas, en ocasiones con el apoyo de líderes de opinión, donde se insiste mucho en algunos resultados y poco en los métodos del ensayo o en las variables de seguridad. Parece muy conveniente entonces que los profesionales sanitarios seamos autónomos en la evaluación de la literatura médica. En la tabla 4 se resume la propuesta de Montori<sup>36</sup> para facilitar la interpretación de los resultados. Estas recomendaciones tienen mucha más relación con el juicio clínico que con el análisis estadístico. Finalmente proponemos dos estrategias para adquirir habilidades en lectura crítica:

1) La consulta de fuentes secundarias de información. Estas publicaciones filtran las principales revistas seleccionando artículos en función de su calidad y relevancia y añadiendo comentarios de los revisores (tabla 5). Estas revistas incluyen además secciones específicas sobre la interpretación de la literatura médica. Afortunadamente cada vez existen más recursos de libre acceso y en castellano, como *Evidence Based Medicine* (<http://ebm.isciii.es>) o bancos de CATs (temas valorados críticamente) como el del Hospital Donostia de San Sebastián (MBE Gipuzkoa), que cuenta con más de 700 artículos resumidos, de los cuales unos 500 son ECA (<http://www.donostiaospitalea.org/MBE/CATS>).

2) El aprendizaje formal en lectura crítica. La autoformación del profesional en lectura crítica de ECA es un proceso progresivo que puede iniciarse respondiendo las preguntas que surgen de la práctica diaria. Este primer paso puede complementarse con programas específicos de formación. Una de las iniciativas más conocidas es el programa CASP (*Critical Appraisal Skills Programme*)<sup>37</sup>. Este movimiento que surgió en el Reino Unido está presente en España a través de CASPe<sup>38</sup> y promueve la formación en lectura crítica para todos los profesionales relacionados con la salud.

**Agradecimientos**

Al grupo MBE de Osatzen (Sociedad Vasca de Medicina Familiar), Comité editorial del INFAC (Boletín Farmacoterapéutico del País Vasco) y en especial a Eulali Mariñelarena, Joserra Aguirrezabala, Carmela Mozo y Josune Iribar por su colaboración en la revisión de la bibliografía.

**Bibliografía**

- Importante.
  - Muy importante.
1. Coumou HC, Meijman FJ. How do primary care physicians seek answers to clinical questions? A literature review. *J Med Libr Assoc.* 2006;94(1):55-60.

**TABLA 5**  
**Selección de recursos electrónicos con resumen de evaluación de ensayos clínicos aleatorizados**

| RECURSO  | DIRECCIÓN   |
|--|---|
| Banco de CATs del grupo MBE-Gipuzkoa                 | <a href="http://www.donostiaospitalea.org/MBE/CATS-basedatos2.asp">http://www.donostiaospitalea.org/MBE/CATS-basedatos2.asp</a> |
| CAT-Crawler  | <a href="http://www.bii.a-star.edu.sg/research/mig/cat_search.asp">http://www.bii.a-star.edu.sg/research/mig/cat_search.asp</a> |
| Bandolera (traducción de <i>Bandolier</i> )          | <a href="http://www.infodoctor.org/bandolera">http://www.infodoctor.org/bandolera</a>   |
| <i>Bandolier</i>                                     | <a href="http://www.jr2.ox.ac.uk/Bandolier/">http://www.jr2.ox.ac.uk/Bandolier/</a>   |
| BestBETs   | <a href="http://www.bestbets.org/">http://www.bestbets.org/</a>   |
| ACP-Journal Club                                     | <a href="http://www.acpjc.org/">http://www.acpjc.org/</a>   |
| Evidence Based Medicine                              | <a href="http://ebm.bmjournals.com/">http://ebm.bmjournals.com/</a>   |
| Evidence Based Medicine, ed. española                | <a href="http://ebm.isciii.es/">http://ebm.isciii.es/</a>   |
| POEMs: <i>Patient Oriented Evidence that Matters</i> | <a href="http://www.infopoems.com/index.cfm">http://www.infopoems.com/index.cfm</a>   |
| Más recursos parecidos en                            | <a href="http://www.infodoctor.org/rafabravo/revistassecun.htm">http://www.infodoctor.org/rafabravo/revistassecun.htm</a>       |



2. Jadad A. Randomised controlled trials. London: BMJ Publishing Group; 1998.
3. • Scott IA, Greenberg PB. Cautionary tales in the clinical interpretation of the therapeutic trial reports. *Intern Med J*. 2005; 35(10):611-21.
4. • Using evidence to guide practice. *MeReC Briefing*. 2005; 30:1-8.
5. • Peiró S. La construcción de la evidence based medicine. *Gestión Clínica y Sanitaria*. 2005;7(4):131-8.
6. • Melander H, hqvist-Rastad J, Meijer G, Beermann B. Evidence based medicine-selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ*. 2003;326(7400):1171-3.
7. • Smith R. Medical journals are an extension of the marketing arm of pharmaceutical companies. *PLoS Med*. 2005; 2(5):e138.
8. • Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ*. 2003;326(7400):1167-70.
9. Is-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA*. 2003; 290(7):921-8.
10. • Evaluating clinical evidence-what do I need to know? *WeMeReC Bulletin*. 2006;May:1-6.
11. • Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. *JAMA*. 1993;270(17):2093-5.
12. Evans A, Kalra L. Are the results of randomized controlled trials on anticoagulation in patients with atrial fibrillation generalizable to clinical practice? *Arch Intern Med*. 2001;161(11):1443-7.
13. Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Pérez A, et al. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. *N Engl J Med*. 1999; 341(10):709-17.
14. Juurlink DN, Mamdani MM, Lee DS, Kopp A, Austin PC, Laupacis A, et al. Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. *N Engl J Med*. 2004;351(6):543-51.
15. Psaty BM, Weiss NS, Furberg CD. Recent trials in hypertension: compelling science or commercial speech? *JAMA*. 2006;295(14):1704-6.
16. Dahlöf B, Sever PS, Poulter NR, Wedel H, Beevers DG, Caulfield M, et al. Prevention of cardiovascular events with an antihypertensive regimen of amlodipine adding perindopril as required versus atenolol adding bendroflumethiazide as required, in the Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm (ASCOT-BPLA): a multicentre randomised controlled trial. *Lancet*. 2005; 366(9489):895-906.
17. Espeland MA, Marcovina SM, Miller V, Wood PD, Wasilaskas C, Sherwin R, et al. Effect of postmenopausal hormone therapy on lipoprotein(a) concentration. PEPI Investigators. *Postmenopausal Estrogen/Progestin Interventions*. *Circulation*. 1998;97(10):979-86.
18. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288(3):321-33.
19. Pitt B, Segal R, Martínez FA, Meurers G, Cowley AJ, Thomas I, et al. Randomised trial of losartan versus captopril in patients over 65 with heart failure (Evaluation of Losartan in the Elderly Study, ELITE). *Lancet*. 1997;349(9054):747-52.
20. Pitt B, Poole-Wilson PA, Segal R, Martínez FA, Dickstein K, Camm AJ, et al. Effect of losartan compared with captopril on mortality in patients with symptomatic heart failure: randomised trial - the Losartan Heart Failure Survival Study ELITE II. *Lancet*. 2000;355(9215):1582-7.
21. • Montori VM, Permyer-Miranda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. *BMJ*. 2005; 330(7491):594-6.
22. Trial of invasive versus medical therapy in elderly patients with chronic symptomatic coronary-artery disease (TIME): a randomised trial. *Lancet*. 2001;358(9286):951-7.
23. Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004;141(10):781-8.
24. Connolly SJ, Gent M, Roberts RS, Dorian P, Roy D, Sheldon RS, et al. Canadian implantable defibrillator study (CIDS): a randomized trial of the implantable cardioverter defibrillator against amiodarone. *Circulation*. 2000;101(11):1297-302.
25. Bokhari F, Newman D, Greene M, Korley V, Mangat I, Dorian P. Long-term comparison of the implantable cardioverter defibrillator versus amiodarone: eleven-year follow-up of a subset of patients in the Canadian Implantable Defibrillator Study (CIDS). *Circulation*. 2004;110(2):112-6.
26. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273(5):408-12.
27. • 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(15):1987-91.
28. Hansson L, Lindholm LH, Niskanen L, Lanke J, Hedner T, Niklason A, et al. Effect of angiotensin-converting-enzyme inhibition compared with conventional therapy on cardiovascular morbidity and mortality in hypertension: the Captopril Prevention Project (CAPPP) randomised trial. *Lancet*. 1999;353(9153):611-6.
29. Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ*. 2002;325(7365):652-4.
30. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW, for the CONSORT Group. Reporting of Noninferiority and Equivalence Randomized Trials: an extension of the CONSORT statement. *JAMA*. 2006;295(10):1152-60.
31. Fox KA, Mehta SR, Peters R, Zhao F, Lakkis N, Gersh BJ, et al. Benefits and risks of the combination of clopidogrel and aspirin in patients undergoing surgical revascularization for non-ST-elevation acute coronary syndrome: the Clopidogrel in Unstable angina to prevent Recurrent ischemic Events (CURE) Trial. *Circulation*. 2004;110(10):1202-8.
32. Bucher HC, Weinbacher M, Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *BMJ*. 1994; 309(6957):761-4.
33. • Hux JE, Naylor CD. Communicating the benefits of chronic preventive therapy: does the format of efficacy data determine patients' acceptance of treatment? *Med Decis Making*. 1995;15(2):152-7.
34. Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet*. 2005;365(9471): 1657-61.
35. Freemantle N. Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we look the crazy aunt in the attic? *BMJ*. 2001;322(7292):989-91.
36. • Montori VM, Jaeschke R, Schunemann HJ, Bhandari M, Brozek JL, Devereaux PJ, et al. Users' guide to detecting misleading claims in clinical research reports. *BMJ*. 2004; 329(7474):1093-6.
37. Critical Appraisal Skills Programme (CASP) and evidence-based practice. Disponible en: <http://www.phru.nhs.uk/casp/casp.htm> [Acceso enero 2007].
38. • Caspe. Programa de habilidades en lectura crítica Española. Disponible en: <http://www.redcaspe.org/homecaspe.asp> [Acceso enero 2007].