



## SUMARIO

*Leyendo entre líneas los ensayos clínicos* ..... 25

Sabemos que la mayoría de nuestros lectores conocen la metodología básica de los ensayos clínicos. En los últimos años se han publicado diferentes propuestas de guiones de revisión (las llamadas "check list") que facilitan la revisión crítica de los ensayos clínicos cuando se aplican en sus dos vertientes: los protocolos de ensayos clínicos para decidir si cumplen los requisitos necesarios para su realización y los ensayos ya publicados para que al lector le resulte más fácil decidir si sus conclusiones son creíbles. Con esta última acepción fue publicado hace muchos años (enero de 1985) un artículo de gran claridad y orientación docente en el volumen 23 del prestigioso "Drug and Therapeutic Bulletin" publicado por la Consumer's Association del Reino Unido. Aunque han pasado muchos años, la metodología del ensayo clínico sigue siendo la misma y, en nuestra opinión, este artículo no ha sido superado en claridad y orientación práctica. Por esta razón lo hemos traducido y reproducido en nuestro boletín con la amable autorización de los editores del DTB.

## LEYENDO ENTRE LÍNEAS LOS ENSAYOS CLÍNICOS

El médico precisa información fiable sobre la eficacia y los riesgos de los tratamientos que aplica. Esta información la proporciona un buen ensayo clínico. Sin embargo, a pesar del rigor creciente de los directores y revisores de las revistas médicas, los ensayos publicados tienen frecuentes defectos de diseño y de análisis que invalidan sus conclusiones y algunas veces confunden al lector (N Eng J Med, 1979; 301: 180, Br Med J, 1977; 1: 85). La mayoría de los médicos tendrán que confiar en revisiones o en literatura de origen comercial, en lugar de en los trabajos originales. Esto puede confundirlos, ya que algunas revisiones pueden basarse, inconscientemente, en ensayos no válidos, aunque parece también que algunos laboratorios lo hacen cuando les conviene.

Muchos médicos leen los artículos de los ensayos que son más relevantes para su práctica. Si lo hacen de manera crítica, pueden decidir si han de aceptar las conclusiones, matizarlas o rechazarlas (Canad Med Assoc J. 1981; 124: 1156). Esto es particularmente importante cuando se leen publicaciones en revistas o suplementos patrocinados por un laboratorio farmacéutico, que pueden no haber estado suficientemente sometidos a revisión (Drug Ther Bull, 1983; 21:24).

### PROBLEMAS DE DISEÑO

Pocas veces un análisis y una interpretación complicada de los resultados pueden salvar un estudio mal diseñado. La revisión puede identificar problemas tras los cuales difícilmente vale la pena leer los resultados y las conclusiones. Si el diseño de los ensayos no está bien descrito, los hallazgos pueden ser sospechosos. La sección de métodos tendría que describir los siguientes puntos.

**Definición de los pacientes.** El artículo tiene que especificar qué tipo de pacientes se incluyeron en el estudio, cuantos fueron excluidos (Br Med J, 1984; 289: 1281) y porqué. Otros detalles, como la edad y el sexo, la procedencia del paciente (sobre todo si era un hospital o una práctica general), los criterios, diagnósticos y gravedad de la enfermedad, los factores de riesgo y otras enfermedades y tratamientos anteriores, ayudan al lector a relacionar los ensayos con su propia práctica. Cuando los pacientes incluidos habían recibido anteriormente uno de los tratamientos sometidos a prueba y no respondieron, se produce una fuente frecuente de sesgo. Los pacientes obtenidos de un grupo seleccionado y no representativo limitan la validez externa del estudio. Así, con frecuencia los ancianos son excluidos de

los ensayos, pero muchas veces se les dan los fármacos probados.

**Definición de los tratamientos.** El artículo tiene que describir cada tratamiento: forma farmacéutica, vía y frecuencia de administración, cómo se escogió la dosis y si fue o no modificada en respuesta a efectos beneficiosos o indeseables. Tiene que hacerse constar otros medicamentos tomados. En los estudios ambulatorios se tiene que evaluar el cumplimiento de la prescripción, como mínimo mediante el recuento de comprimidos.

**Métodos de evaluación.** El lector tiene que poder decidir si el método de evaluación de la respuesta al tratamiento es relevante para sus pacientes. Cada medida clínica ha de estar definida de manera precisa y siempre que sea posible, ha de ser objetiva y reproducible. Frecuentemente los sistemas de puntuación arbitraria no cumplen estos requisitos y si se emplean, hay que hacerlo a ciegas, para evitar sesgos. Si es posible las medidas deberán ser realizadas por un único observador (convenientemente entrenado), o deberán estar explícitamente normalizadas, sobre todo en los ensayos multicéntricos. El protocolo ha de permitir una recogida correcta de los efectos no deseados, tanto los referidos espontáneamente por el paciente, como los obtenidos mediante cuestionario específico.

**Grupo de control.** Los ensayos clínicos no controlados, en los cuales todos los pacientes reciben el mismo tratamiento, pocas veces dan información útil sobre la eficacia, ya que no pueden demostrar si el tratamiento ha producido algún cambio, excepto cuando el desenlace de la enfermedad es rápido y previsible. Pueden ser útiles para detectar la toxicidad. Para evaluar los méritos de un fármaco, este tiene que ser comparado con el tratamiento habitual o bien con un placebo. Los ensayos clínicos controlados con placebo son adecuados al principio del desarrollo de un fármaco o cuando no hay ningún tratamiento de eficacia comprobada. Cuando la mejoría espontánea es frecuente, o también cuando los dos tratamientos activos consiguen resultados semejantes (y por tanto puede ser imposible decidir si son eficaces o ineficaces), también puede haber un grupo de control con placebo, además del grupo de control sometido a un tratamiento activo. Los ensayos controlados con placebo, son más frecuentes, pero a veces incorrectos, porque pueden dar diferencias impresionantes y estadísticamente significativas en un número reducido de pacientes. Raramente ayudan al clínico que quiere saber si el nuevo fármaco es mejor que el tratamiento anterior, y en muchas enfermedades puede considerarse que no son éticos.

En los ensayos en los que el grupo de control está sometido a un tratamiento activo, éste tiene que ser el mejor disponible, ya que en caso contrario el ensayo tampoco sería de mucha utilidad para el clínico. Desafortunadamente, con frecuencia los labo-

ratorios patrocinan ensayos en los cuales el tratamiento de comparación no es el adecuado, y así las conclusiones resultan distorsionadas. La elección de la dosis también es objeto de abusos: en los ensayos en que se compara la eficacia de fármacos con acciones similares, se han de utilizar dosis que puedan dar lugar a una incidencia igual de efectos no deseados, y en los ensayos en los que se compare la toxicidad se han de utilizar dosis de la misma efectividad.

**Distribución al azar ('randomization').** Los pacientes de cada grupo tienen que ser semejantes en sus características relevantes antes de empezar el tratamiento. La única manera fiable de conseguirlo es asignando a cada paciente al azar a uno u otro grupo. Las comparaciones históricas del tratamiento probado en pacientes tratados en un período anterior no son válidas. Los ensayos no controlados o no randomizados pueden distorsionar mucho la eficacia de un nuevo fármaco; son cada día menos frecuentes, pero aún aparecen en documentos promocionales y en publicaciones patrocinadas.

**Problemas en la distribución al azar.** En muchos ensayos no se hace una verdadera distribución al azar. El paciente o el investigador pueden rehusar el tratamiento asignado. Así, por ejemplo, en la asignación alterna, si el investigador conoce qué tratamiento es el siguiente y selecciona un caso 'adecuado', los pacientes pueden ser asignados de manera no aleatoria. Hay otros métodos poco eficaces que pueden dar lugar a una mala distribución al azar, por ejemplo, el empleo de la fecha de nacimiento del paciente. Estos defectos se pueden sospechar si los grupos de tratamiento formados presentan (involuntariamente) un número diferente de pacientes o difieren en características básicas (Br Med J, 1984; 289: 336). Cuando no se comparan las características relevantes de cada grupo se puede, estar ocultando algún sesgo.

Los diseños cruzados, en los que cada paciente recibe dos o más tratamientos, uno después de otro, pueden ser eficientes para comparar tratamientos breves en enfermedades crónicas, en un curso relativamente constante. Sin embargo, los hallazgos a corto plazo no son necesariamente aplicables a las necesidades que los pacientes tienen a largo plazo. El orden de los tratamientos ha de ser randomizado y tiene que haber un período de 'limpieza' entre uno y otro tratamiento, a fin de impedir una sobreposición de efectos. Es útil incluir también un período de observación 'de entrada', para demostrar que la enfermedad tiene un curso estable.

**Ciego simple, doble ciego.** Cuando el paciente o el investigador conocen qué tratamiento se está administrando, pueden sesgar la respuesta y su medida objetiva. En los ensayos a doble ciego, en los que únicamente se revela el tratamiento cuando ya se han evaluado todos los pacientes, se evita este sesgo. Se tiene que describir la técnica del doble cie-

go, así como el preparado del placebo, si se utiliza. Si los dos fármacos difieren en su aspecto o en su frecuencia de dosificación, se pueden administrar las dos formas farmacéuticas a los dos grupos de pacientes: una contiene el tratamiento de prueba y el otro el de referencia. A veces el diseño a doble ciego es imposible (por ejemplo en ensayos sobre técnicas quirúrgicas) y a veces se afirma que este fue el diseño del estudio, pero eso es materialmente imposible (por ejemplo cuando un fármaco tiene un efecto o un sabor característico). En estos casos quizá hay que aceptar un diseño parcialmente ciego, con evaluaciones independientes de la respuesta y de los efectos no deseados, con las naturales precauciones.

**Tamaño del ensayo.** La mayoría de los ensayos incluyen demasiado pocos pacientes para obtener resultados fiables (N Eng J Med, 1978; 299: 690). El número de pacientes necesarios se puede calcular a partir de la proporción de los que se espera que respondan y a partir de la diferencia de eficacias que se considera clínicamente relevante (Br Med J, 1980; 281: 1336). Demostrar que un fármaco es más eficaz que el placebo necesita generalmente muchos menos pacientes que demostrar la superioridad sobre otro fármaco activo. Demostrar que dos fármacos son igualmente eficaces va a requerir un ensayo con muchos pacientes. Si no se describe como se tomó la decisión de dar por finalizado el ensayo, puede ser que se hubiesen realizado múltiples análisis intermedios de los datos y que el estudio se terminó cuando los resultados eran más llamativos. Esto exagera el beneficio aparente, y no es válido (excepto si se utiliza un diseño "secuencial").

**Sesgo de publicación.** Los ensayos pequeños afectan seriamente la evaluación de las investigaciones publicadas. Los ensayos pequeños con resultados negativos, pocas veces son publicados en revistas médicas importantes, de manera que sólo reciben publicidad los que presentan beneficios: esto constituye una selección optimista de la masa mucho más grande de ensayos pequeños negativos no publicados. Por tanto, cabe desconfiar de un ensayo pequeño que anuncie un gran avance: incluso si está bien diseñado y analizado, su contribución puede haber resultado exagerada por el sesgo de la publicación, favorable a los estudios positivos.

## ANÁLISIS

Si hay graves defectos en el diseño, son pocas las ocasiones en que vale la pena seguir leyendo, ya que puede ser difícil evitar la atracción que ejercen los resultados positivos, aunque sean no válidos. Los defectos menores pueden no invalidar el ensayo, pero precisan una reinterpretación de los resultados y solamente permiten extraer conclusiones limitadas pero útiles. Un ensayo bien diseñado también

puede ser invalidado por un análisis incorrecto de los resultados.

Se han de examinar los resultados y su interpretación, con tal de encontrar lo que se quiere saber, y no lo que el investigador quiere decir. El artículo tiene que permitir que los lectores interpreten ellos mismos los resultados. La publicación e interpretación abusiva únicamente de una parte de los resultados, así como el mal uso de las técnicas estadísticas, puede dar lugar a una conclusión inválida y exagerada.

**No se explica que pasó con todos los pacientes.** Hay que explicar todas las retiradas de pacientes. El ensayo puede ser abandonado por muchas razones; generalmente por causa de efectos no deseados o por empeoramiento. Puede haber más pérdidas en un grupo que en el otro. Algunos autores excluyen estas pérdidas de los resultados. Esto distorsiona las conclusiones, sobre todo cuando la razón de la retirada podría tener que ver con la eficacia o la toxicidad, y puede ocultar efectos adversos importantes. Las publicaciones que no explican que ha pasado con todos los enfermos que ya habían sido asignados a uno u otro grupo de tratamiento pueden ser sesgadas. En los ensayos en los que de todas maneras se puede registrar lo que pasó con los pacientes retirados (por ejemplo los estudios de supervivencia), la práctica aceptada es incluir los retirados en un análisis estadístico adicional ('análisis por intención de tratamiento'), de esta manera se evita el optimismo inherente a los análisis 'solamente de cumplidores'. El primer análisis revela el valor del fármaco en la práctica clínica, la segunda sus efectos específicos.

**Falta de datos individuales.** Con frecuencia los resultados se presentan como valores medios con sus errores estándar. Aunque esto puede constituir una base válida para la comparación con una prueba de significación, ignora la naturaleza y la variedad de las respuestas individuales de cada paciente; alguno puede que no respondiese de ninguna manera. Por lo tanto, es útil presentar una tabla o gráfica en la que se resume la respuesta de cada paciente. La presentación de los índices de respuesta exclusivamente como porcentajes, en vez de presentarlos en números absolutos de pacientes, también puede causar confusión, sobre todo si la muestra es pequeña.

**Obsesión con las pruebas de significación.** Las pruebas de significación estadística evalúan únicamente la probabilidad que la diferencia observada entre los grupos pudo producirse al azar. La expresión ' $p < 0,05$ ' no indica que el tratamiento sea mejor, tan solo quiere decir que la diferencia observada se encontraría por casualidad en menos de uno de cada veinte ensayos en los que no existiesen diferencias reales entre los tratamientos. Cuantas más pruebas de significación se realizan en un ensayo, más probabilidades hay de que aparezca algún falso positivo

por azar (error de tipo I), a no ser que se ajuste por 'contraste múltiple'. El sesgo de publicación en favor de los ensayos 'positivos' mencionado anteriormente, también incrementa el riesgo de publicar falsos resultados positivos ('p' quiere decir probabilidad, no publicabilidad). De manera análoga, 'ns' (no significativo) no quiere decir que no hay diferencias entre ambos tratamientos; únicamente quiere decir que las diferencias existentes no son lo suficientemente grandes como para ser detectadas con el poder estadístico del estudio. Con frecuencia se utilizan los ensayos pequeños para 'demostrar' que dos fármacos no difieren entre ellos, pero pocas veces son capaces: en realidad pueden esconder una diferencia importante y real (N Eng J Med, 1978; 299:690).

**Importancia de la medida.** Una diferencia 'significativa' puede ser demasiado pequeña para ser clínicamente relevante y un resultado no significativo puede esconder una diferencia importante que precisaría de un ensayo con más pacientes para poder ser demostrada (error de tipo II). Sólomente los límites de confianza pueden dar una idea de la cuantía del beneficio del tratamiento, pero este sencillo cálculo pocas veces está incluido en el artículo (o es solicitado por los editores de las revistas médicas). La afirmación de que un tratamiento produce, por ejemplo, una mejoría del 50% no es una ayuda si no se conoce el grado de incertidumbre alrededor de esta cifra, quiere decir, los límites de confianza. Los ensayos clínicos pequeños no pueden cuantificar de manera precisa la magnitud del tratamiento con precisión, incluso si producen resultados 'significativos'.

**"Selección" de resultados.** Algunos autores analizan los datos obtenidos de muchas maneras y después publican sólomente aquellos resultados que dan diferencias 'significativas' o únicamente insisten en estos. Así también aumenta el riesgo de falsos positivos, que no se pueden distinguir de los 'verdaderos' positivos. Esta selección de resultados puede ser difícil de detectar para el lector. Así, por ejemplo, si no hay diferencias globales, cada grupo puede ser dividido en subgrupos aparentemente relevantes siendo publicados únicamente en el caso de presentar diferencias significativas. Esta estratificación sería más aceptable si previamente se describiese en el protocolo original. Las principales características medidas deberían definirse en el momento de diseñar el ensayo, puesto que en caso contrario los autores pueden escoger retrospectivamente sobre qué medidas insistir en función de su 'significación' estadística. Esto no impide que se presente material adicional, pero hay que declarar todas las hipótesis probadas y la hipótesis inicial debe ser la base de la conclusión de los ensayos. Si la respuesta va a ser repetidamente medida durante un período de tiempo, hay que emplear todas las medidas en un análisis único y apropiado, en vez de hacer pruebas estadísticas sobre cada período por separado y de presentar tan solo las diferencias 'significativas'.

Inversamente, hay que comprobar que las comparaciones que serían obviamente relevantes y factibles no se omitan. Ya que pudieron omitirse bien porque no eran significativas, bien porque no coincidían con la finalidad buscada. Si la conclusión del ensayo se basa en una diferencia significativa acompañada de muchas comparaciones no significativas, hay que sospechar un falso resultado positivo.

**Métodos estadísticos inapropiados.** El conocimiento de la estadística básica puede permitir al lector la detección de pruebas estadísticas inapropiadas (por ejemplo pruebas de la 't' en ensayos pequeños con datos sesgados, cuando serían más apropiadas las pruebas no paramétricas). Si parece que sería aplicable un método más sencillo y bien conocido, hay que explicar la necesidad y el significado de un análisis complejo. Los ensayos bien diseñados producen resultados que suelen ser sencillos de interpretar.

**Gráficas y tablas distorsionadas.** La presentación de los datos obtenidos en ilustraciones pueden exagerar las diferencias. En una gráfica una escala ampliada o partida puede hacer que una diferencia sin importancia biológica parezca enorme. Las gráficas y los histogramas de respuestas medias a cada tratamiento que no incluyan los errores estándar (o los límites de confianza), pueden hacer que una diferencia no significativa parezca impresionante. Estas técnicas son muy comunes en folletos publicitarios, y no son raras en las revistas médicas.

**Falta de beneficios reales.** El lector tiene que comprobar que el tipo de respuesta demostrada y su magnitud es lo que precisa el paciente. Los ensayos que únicamente demuestran el efecto del tratamiento durante un breve período de tiempo o que miden parámetros que sólomente tienen una relación indirecta con el estado general del paciente, son poco útiles para decidir la elección de un tratamiento. Puede ser que el clínico necesite traducir los resultados de variables más 'duras', necesarias para evitar los sesgos, a una medida más humana del beneficio para la calidad de vida del paciente. La publicación tiene que comparar los efectos no deseados y beneficiosos del tratamiento, aunque su incidencia únicamente podrá ser comparada, de manera precisa, en ensayos con muchos pacientes y mayor tiempo de seguimiento.

**Conclusiones injustificables.** Aunque los datos estén correctamente analizados y presentados, el resumen y las conclusiones pueden reflejar más el entusiasmo de los autores que lo que realmente es justificable a partir de los resultados del ensayo y de otros datos. La conclusión tiene que presentar una visión equilibrada de los datos disponibles y no ha de dejar de lado los resultados que no coincidan con el comprensible optimismo de los autores y del patrocinador.