

Controversias metodológicas actuales sobre ensayos clínicos controlados y meta-análisis

V. Barrio

Unidad de Nefrología. Area de Medicina. Fundación Hospital Alcorcón.

INTRODUCCION

A la hora de abordar la lectura crítica de la literatura en medicina que ha alcanzado enormes dimensiones por su imparable crecimiento en las últimas décadas, tiene especial relevancia el método estadístico o en su sentido más amplio la epidemiología clínica para servir como guía y permitir establecer criterios objetivos de clasificación de los estudios en función de la valoración crítica de su metodología y de la interpretación de sus resultados. Dada la extensión y complejidad del problema, no se pretende en la presente revisión abordar temas como la importancia del tipo de diseño del estudio, la utilización de test estadísticos apropiados en cada caso o la creciente utilización en la literatura médica actual de los intervalos de confianza en detrimento del valor de «p» o de los contrastes de hipótesis¹⁻³, sino revisar brevemente algunas de las últimas controversias metodológicas publicadas en la literatura médica sobre los ensayos clínicos controlados y el meta-análisis, ya que estos dos tipos de estudios conforman la piedra angular sobre la que se ha desarrollado la corriente actual de la medicina basada en la evidencia.

ENSAYOS CLINICOS CONTROLADOS

El ensayo clínico controlado es el paradigma del estudio clínico experimental para demostrar la causalidad y eficacia de cualquier intervención, ya sea diagnóstica o terapéutica, principalmente porque el proceso de aleatorización del tratamiento permite equilibrar los grupos no sólo para factores pronósticos o de confusión conocidos, sino también para aquellos desconocidos o imposibles de controlar, evitando el mayor problema de la in-

vestigación clínica que es el sesgo de selección. Por tanto, no es de extrañar que desde la primera publicación en el año 1948 de un estudio de investigación clínica con este diseño demostrando la eficacia del tratamiento con estreptomycin en la tuberculosis pulmonar⁴, hayan proliferado este tipo de estudios al constituir la mejor evidencia disponible del efecto de las intervenciones médicas.

Sin entrar a discutir cuestiones metodológicas de diseño, desarrollo y análisis de los ensayos clínicos que conforman la validez interna del estudio y que fueron magistralmente expuestos en dos publicaciones^{5,6} hace más de veinte años por una pleyade de epidemiólogos y estadísticos considerados por muchos como los pioneros de la epidemiología clínica moderna, ni tampoco de la interpretación de sus resultados como sería el impacto del efecto y la precisión de la estimación, ya que este aspecto ha sido revisado en fecha reciente de forma exhaustiva aunque más pragmática por el grupo promotor de la medicina basada en la evidencia^{7,8}, en las siguientes líneas sólo se pretende ilustrar algunos aspectos metodológicos de los ensayos clínicos controlados como son la utilización del placebo, el análisis de subgrupos, el papel de los comités de ética e investigación clínica o las normas de publicación, que han suscitado cierta controversia en los últimos años y siguen estando de actualidad.

UTILIZACION DEL PLACEBO

Cada vez está más controvertida la utilización del placebo en los ensayos clínicos controlados, ya que no resulta ético si existe alguna intervención que previamente haya demostrado eficacia en esas circunstancias^{9,10}. Por lo tanto, se podría esperar que a medida que avanza el conocimiento médico, debería reducirse el número de ensayos con placebo en el grupo control. Como corolario de este hecho, también deberían ser cada vez más frecuentes los ensayos de equivalencia terapéutica, aunque esto no parece ocurrir en la práctica ya que muchos en-

Correspondencia: Dr. V. Barrio
Fundación Hospital Alcorcón
C/ Budapest, 1
28922 Alcorcón. Madrid

sayos que deberían ajustarse a las modificaciones metodológicas particulares de los ensayos de equivalencia, no las cumplen al plantearse como ensayos de eficacia terapéutica. Las principales modificaciones conceptuales de un ensayo de equivalencia terapéutica se podrían resumir¹¹ en que el tamaño del ensayo se debe calcular de acuerdo con la hipótesis nula de no equivalencia en lugar de la habitual ausencia de diferencia entre los diferentes tratamientos, que para obtener una conclusión afirmativa, el intervalo de confianza de la diferencia entre las intervenciones debe estar comprendido en el rango de equivalencia predefinido, que se deben realizar los análisis tanto por intención de tratar como por cumplimiento de protocolo y que se deben extremar los controles para asegurar la validez interna del ensayo y poder demostrar no sólo la equivalencia de los tratamientos sino también su eficacia.

Incluso se cuestiona recientemente, tanto por cuestiones éticas ya que en muchos consentimientos informados no se menciona explícitamente, como por su pobre utilidad desde el punto de vista estadístico, la utilización ciega del placebo en el período de lavado previo a la asignación aleatoria del tratamiento en los ensayos clínicos controlados¹². Es una práctica muy frecuente que habitualmente se justifica por la necesidad de excluir del estudio a los sujetos con mal cumplimiento al tratamiento, a los respondedores a placebo o a aquellos sujetos con mayor variabilidad de respuesta. Aunque en este supuesto se obtendría una mayor homogeneidad de la población de estudio, por este mismo motivo limitaría la validez externa y su aplicación a la práctica clínica diaria. En otros casos, se utiliza para lavar los efectos de tratamientos previos y tener un período basal de referencia con el que comparar el efecto de la intervención objeto de estudio. Las objeciones a su empleo en estos casos se derivan en que desvirtúan el objetivo primordial del estudio, ya que lo realmente relevante no es medir el curso natural del tratamiento activo sino su efecto causal.

ANALISIS DE SUBGRUPOS

No es infrecuente encontrar en las publicaciones de ensayos clínicos controlados, hasta el 50% en algunas revisiones¹³, análisis de subgrupos que comparan la respuesta al tratamiento en diferentes categorías de pacientes o que destacan entre sus resultados, los obtenidos al analizar objetivos secundarios, incluso no especificados formalmente en el protocolo inicial de estudio. Como ejemplo, recordar la

sorprendente reducción de mortalidad global en pacientes mayores de 65 años tratados con un bloqueante de los receptores de angiotensina II frente a un inhibidor de la enzima de conversión de la angiotensina encontrada en un ensayo recientemente publicado¹⁴, cuando el diseño perseguía comprobar la tolerabilidad de ambos fármacos expresada como deterioro de función renal que se estimaba por el aumento de las cifras de creatinina sérica. La importancia de dicho hallazgo ha obligado a la puesta en marcha de un nuevo ensayo clínico, que ya ha comenzado, para confirmar la veracidad del resultado. Los hallazgos de los análisis de subgrupos pueden tener importantes consecuencias negativas, tanto en el sentido de alcanzar una conclusión falsamente negativa, es decir, privar de un tratamiento efectivo a una determinada categoría de pacientes, como en el sentido de alcanzar una conclusión falsamente positiva lo que supondría tratar con una intervención ineficaz e incluso potencialmente peligrosa al no estar exenta de efectos adversos a otro subgrupo de pacientes. Sin embargo es fundamental reconocer que pueden y suelen existir características diferenciales tanto de los pacientes (edad, sexo, severidad de enfermedad subyacente, grado y severidad de comorbilidad asociada) como de los tratamientos (cronología, dosis), que pueden ejercer una importante modificación del efecto de la intervención considerada y modular el grado de respuesta. A continuación se indican algunas normas propuestas para decidir si las aparentes diferencias en la respuesta de diferentes subgrupos son reales, de acuerdo con criterios que estiman el potencial de error aleatorio y sistemático, la consistencia de la evidencia y su verosimilitud biológica^{15,16}. En primer lugar, cuanto mayor sea la diferencia entre el efecto global y el observado en un subgrupo particular, más probable es que la diferencia sea real y que tenga relevancia clínica. En segundo lugar se debe considerar la significación estadística de la diferencia sin perder de vista la menor potencia del test al reducir el tamaño de muestra y por tanto la mayor probabilidad de cometer un error de tipo II, es decir de concluir falsamente que no existen diferencias, aunque existen técnicas estadísticas convencionales¹⁷ que permiten realizar formalmente dichos análisis. Los siguientes criterios se refieren al llamado análisis «post-hoc», es decir el realizado *a posteriori*, en cuyo caso debe ser considerado como generador de hipótesis para posteriores estudios más que como comprobante de hipótesis y a la cantidad de análisis de subgrupos realizados, lo que aumentaría considerablemente si el número es alto, la probabilidad de cometer un error de tipo I, es decir de alcanzar conclusiones falsamente positivas a menos de

corregir el valor de «p» para comparaciones múltiples. A continuación, se debe comprobar si las inferencias se basan en diferencias encontradas en el propio estudio o si se derivan de comparaciones entre diferentes estudios, en cuyo caso pierden gran parte de su valor, ya que incluso el proceso de aleatorización no asegura la comparabilidad de los grupos en los diferentes estudios. Finalmente, se debe considerar la consistencia de los resultados, es decir la replicación de una interacción o modificación del efecto en estudios independientes y no sesgados así como el grado de verosimilitud biológica que habitualmente proviene de tres tipos de evidencia indirecta como son la existencia de estudios en diferentes poblaciones, incluyendo experimentación animal, que alcancen las mismas conclusiones, la observación de interacciones para intervenciones similares y los resultados de estudios similares pero que analizan variables o resultados intermedios.

VALIDEZ EXTERNA Y GENERALIZACION DE RESULTADOS

La aplicación de los resultados de la investigación clínica, en particular de los ensayos clínicos controlados a la práctica diaria de la medicina, está íntimamente ligada a los conceptos de generalización y validez externa de los estudios, aunque se puede considerar que tiene una amplitud mayor al abarcar otras cuestiones relacionadas con el impacto global del tratamiento en cada paciente en particular. No se debe basar exclusivamente en los criterios de inclusión y exclusión utilizados para seleccionar a la población del ensayo clínico, ya que estos habitualmente se definen para maximizar el poder estadístico del estudio y minimizar los riesgos, sino que debe tener en cuenta factores fisiopatológicos, epidemiológicos e incluso sociales y económicos que pueden modificar la reducción de riesgo esperada tanto como el grado de daño o efectos adversos de la intervención considerada¹⁸. Con respecto a los factores fisiopatológicos, se debe valorar si existen diferencias tanto en el proceso patológico (diferente respuesta biológica, resistencia a fármacos) como en el propio huésped (respuesta inmune, biodisponibilidad, farmacocinética) capaces de modificar en un sentido u otro la respuesta al tratamiento, lo que alteraría la relación riesgo/beneficio y por tanto la actitud terapéutica final. También debería tomarse en cuenta la potencial influencia del medio sociocultural y de las condiciones económicas sobre la efectividad del tratamiento propuesto por las diferencias en la adherencia al tratamiento por parte de los sujetos así como en su disponibilidad y correcta dis-

pensación por parte de los profesionales sanitarios. Sin embargo, a mi juicio los puntos claves que modulan la generalización de los resultados de un ensayo y su aplicación práctica, son los llamados factores epidemiológicos como la existencia de comorbilidad que puede alterar significativamente los potenciales riesgos y/o beneficios del tratamiento y la premisa habitual de que la reducción del riesgo relativo por la intervención se mantiene constante entre los diferentes subgrupos de pacientes, condición que no siempre se cumple¹⁹. Estos problemas han sido revisado en profundidad recientemente²⁰, recomendándose que en primer lugar se debe comprobar, de forma gráfica y/o estadística²¹, que la reducción del riesgo relativo se mantiene constante en los diferentes estratos valorando los factores que pueden modificarla (intensidad, tiempo y modo de administración de la intervención), en segundo lugar, predecir el nivel de riesgo particular del paciente, habitualmente derivado de una predicción de riesgo multivariable que suele provenir de estudios longitudinales prospectivos o incluso de los propios grupos controles de los ensayos clínicos, para finalmente poder estimar el beneficio y riesgo de la intervención en cada caso particular.

COMPROMISO DE MEJORA DE CALIDAD DE PUBLICACION DE ENSAYOS CLINICOS CONTROLADOS

Por su importante y creciente impacto en la práctica diaria y para que la comunidad científica pueda juzgar el rigor metodológico y validez de los resultados publicados, un gran número de publicaciones médicas entre las que se encuentran varias revistas de amplia difusión y alto índice de impacto como JAMA, Ann Intern Med, BMJ y Lancet, entre muchas otras, han adoptado recientemente, dadas las deficiencias en la comunicación de los protocolos de estudio y de los resultados, señaladas de forma repetida desde hace más de veinte años^{13,22,23}, un código de requerimientos mínimos para la publicación de ensayos clínicos controlados denominado CONSORT²⁴, acrónimo de «Consolidated Standards of Reporting Trials», que incluye una guía con los 21 ítems que debe incluir la publicación y un diagrama de flujo que ilustra el desarrollo del ensayo. La guía incluye los puntos que describen prospectivamente entre otros la hipótesis objeto de estudio, los objetivos clínicos y análisis previstos de subgrupos o corregidos por covariables, el protocolo de estudio con la población, criterios de inclusión y exclusión, intervenciones, medidas de resultado primarias y secundarias, diferencias mínimas consideradas clínica-

mente relevantes, cálculo del tamaño muestral, análisis estadístico, criterios de cierre prematuro del ensayo, método y unidad de aleatorización, método de enmascaramiento, así como un diagrama de flujo que muestra el reclutamiento y seguimiento de los sujetos a lo largo del tiempo con pérdidas y abandonos. La adopción de este esquema de comunicar resultados permitirá sin duda, la evaluación rigurosa de la validez interna y externa de los ensayos clínicos. En este sentido, señalar que la Agencia Europea para la Evaluación de los Productos Medicinales (EMA) ha dictado unas excelentes guías sobre la estructura y el contenido de informes sobre ensayos clínicos controlados (CPMP/ICH/137/95) y sobre buena práctica clínica (CPMP/ICH/135/95) que se encuentran disponibles en internet²⁵.

PAPEL DE LOS COMITES DE ETICA E INVESTIGACION CLINICA

Redundando sobre las ideas expuestas en las líneas precedentes, algunos autores incluso reclaman una mayor participación de los comités de ética e investigación clínica, que son los responsables de aprobar la realización de los ensayos clínicos controlados y velar por su correcto desarrollo tanto desde el punto de vista metodológico como ético, para mejorar su calidad científica así como para asegurar su posterior difusión. Así se ha recomendado recientemente²⁶ que los comités de ética e investigación clínica deberían solicitar una revisión sistemática, o meta-análisis, del tema objeto de estudio antes de aprobar el ensayo con lo que se evitaría la repetición superflua de estudios y privar de un tratamiento efectivo a los potenciales participantes en el estudio²⁷, que un resumen de los resultados de dicho análisis con los riesgos y beneficios esperables de la intervención fuera entregado a los potenciales participantes en el mismo, requerir como condición para su aprobación que todo ensayo clínico fuera registrado una vez iniciado, como ya se está haciendo en Gran Bretaña²⁸, lo que evitaría duplicar esfuerzos y permitiría controlar el sesgo de publicación²⁹, que existiera un compromiso formal por parte de los investigadores de hacer públicos los resultados y finalmente permitir la auditoría de la publicación de los resultados de los ensayos previamente aprobados por dicho comité.

META-ANALISIS

El meta-análisis es un término y técnica provenientes de la literatura psicológica^{30,31} que ha sido

definido como el análisis estadístico que se realiza sobre el conjunto de resultados de muchos estudios empíricos con el propósito de integrar sus hallazgos. Representa una buena alternativa a las tradicionales revisiones bibliográficas por su carácter cuantitativo y la exigencia de rigor metodológico en todas las fases de su realización. Aunque en la recopilación de artículos que forman un meta-análisis se incluyan estudios experimentales y/o estudios de observación o transversales, el meta-análisis en sí mismo es un estudio observacional, con las limitaciones propias de este tipo de diseño³². Los principales objetivos del meta-análisis son aumentar la potencia estadística al integrar los resultados procedentes de múltiples estudios para objetivos primarios o secundarios, y por tanto mejorar la fiabilidad y precisión de la estimación de la magnitud de la intervención objeto de estudio, resolver o intentar explicar discrepancias entre resultados dispares de estudios con planteamientos similares, y por último, intentar resolver cuestiones no especificadas *a priori* en los estudios primarios como pueden ser análisis de subgrupos, así como sentar las bases de las futuras líneas de investigación más prometedoras.

A pesar del crecimiento exponencial de los meta-análisis desde finales de los años 70 en la literatura clínica médica, muchas voces críticas han acompañado su progresiva generalización y difusión, señalando las principales limitaciones y sesgos inherentes a este tipo de estudios. A continuación, sin entrar a exponer la metodología del meta-análisis revisada profusamente en los últimos años por múltiples autores³³⁻³⁷ o su interpretación crítica y aplicación a la práctica diaria³⁸, sólo se pretende ilustrar brevemente algunas de las últimas controversias metodológicas sobre este tema así como las diferentes soluciones propuestas.

META-ANALISIS CUMULATIVO Y MONITORIZACION SECUENCIAL

Con objeto de evitar los sesgos más comunes inherentes a este tipo de estudios, se ha recomendado en los últimos años la adopción de unas normas estrictas para el diseño, desarrollo, análisis e interpretación de los meta-análisis^{39,40} cuyos puntos más relevantes se resumen a continuación. En primer lugar y considerando únicamente estándares cualitativos, se ha sugerido que todo meta-análisis debería seguir un protocolo definido *a priori* que incluyera la(s) cuestión(es) objeto de estudio, que identificara, idealmente de forma prospectiva mediante registros, los ensayos clínicos que deberían ser incluidos, con detalles sobre la población y subgrupos de

especial trascendencia, que definiera los tratamientos o intervenciones así como los eventos que marcan el éxito o fracaso de la intervención, que comprobara la integridad de la aleatorización, y finalmente que basara su análisis en intención de tratar y sus conclusiones en la totalidad de la información disponible al incluir todos los ensayos sobre el tema, todos los pacientes randomizados y todos los eventos relevantes recogidos en cada ensayo. Con respecto a los estándares cuantitativos, se ha sugerido recientemente³⁹ que todo meta-análisis debería calcular, al igual que todo ensayo clínico controlado incluye el cálculo del tamaño de muestra, el tamaño óptimo de información, definido como la mínima cantidad de información requerida en una revisión colectiva para alcanzar conclusiones válidas sobre el efecto de una intervención, teniendo en cuenta la tasa de eventos en la enfermedad objeto de estudio y el mínimo efecto del tratamiento que se considere biológicamente plausible y clínicamente relevante. Esto conduciría a establecer, de forma similar a como ocurre con los análisis intermedios que se realizan antes de que finalice el período teórico de seguimiento de los ensayos clínicos controlados, una monitorización secuencial con puntos de corte pre-establecidos para los meta-análisis cumulativos, que consisten en la puesta al día de forma periódica del meta-análisis sobre algún tema cuando se publica un nuevo ensayo, con objeto de identificar lo antes posible el potencial beneficio de una intervención terapéutica. La utilización de esta estrategia de calcular el tamaño óptimo de información y la ponderación de los datos acumulados mediante técnicas estadísticas formales de monitorización secuencial, evitaría las conclusiones prematuras sobre el beneficio del tratamiento basado en meta-análisis de pequeños ensayos y proporcionaría un regla para decidir cuando se ha obtenido una evidencia concluyente.

DISCREPANCIAS ENTRE META-ANÁLISIS DE ENSAYOS DE PEQUEÑO TAMAÑO Y POSTERIORES ENSAYOS CON GRAN TAMAÑO DE MUESTRA

En los últimos años se ha señalado con frecuencia la existencia de importantes discrepancias entre los resultados obtenidos de meta-análisis de ensayos clínicos de pequeño tamaño y los derivados de posteriores ensayos con gran tamaño de muestra o gran poder estadístico para detectar diferencias entre los grupos con tratamiento activo frente al control⁴¹⁻⁴³, también conocidos como megaensayos. En la mayoría de los casos, el megaensayo posterior no confir-

ma el beneficio que mostraba el meta-análisis de los pequeños ensayos iniciales y este sesgo puede incluso predecirse por la asimetría de un gráfico de «embudo» que representa la estimación del efecto de cada ensayo frente al tamaño de la muestra⁴⁴. Como ejemplo, citar la ausencia de efecto o el efecto marginal de dudosa relevancia clínica, de la aspirina a dosis bajas en la prevención de la preeclampsia en los últimos ensayos publicados que incluyen a un alto número de participantes⁴⁵⁻⁴⁷, cuando los pequeños ensayos iniciales de finales de la década de los 80, resumidos en un meta-análisis publicado en el año 91⁴⁸ mostraba una importante reducción del riesgo de desarrollo de preeclampsia (riesgo relativo 0,35, intervalo de confianza [IC] 95%, 0,22-0,55) con el tratamiento preventivo de aspirina a dosis bajas. Sin embargo, de acuerdo con los datos actuales, es necesario tratar a 100 mujeres embarazadas para prevenir un caso de preeclampsia, o incluso restringiendo el tratamiento a aquellas mujeres de alto riesgo, como son las que presentan diabetes mellitus pregestacional, hipertensión crónica, gestación múltiple o preeclampsia en gestaciones previas y cuya incidencia de preeclampsia alcanza hasta el 20%, esta cifra sólo se reduce a tratar a 38 mujeres para prevenir un solo caso de preeclampsia⁴⁷. Por tanto, no solo la diferencia de riesgo del grupo control que se ha invocado como una de las principales razones que podrían explicar las diferencias encontradas²¹ y que en este caso particular era incluso menor en los pequeños ensayos iniciales, ya que alcanzaba un 17,6% de incidencia agrupada de preeclampsia, es capaz de explicar estas diferencias, sino que otras razones parecen ser las responsables de la heterogeneidad entre dichos ensayos.

Se podría resumir diciendo que en función del método utilizado para definir una discrepancia, esta aparece entre un 10 y 25% de los casos, con valores predictivos positivo y negativo cercanos al 33%, aunque esta variabilidad se reduce a la mitad si se utiliza para estimar el efecto agrupado el modelo estadístico de efectos aleatorios. Además, el sesgo de publicación o diferencias en la población o en los protocolos de estudio se han sugerido como explicaciones plausibles para justificar en la mayoría de los casos estas diferencias^{42,43}. El sesgo de publicación se refiere a que tienen menor probabilidad de publicación los ensayos clínicos con resultados negativos²⁹, lo que no ha mejorado mucho a pesar del compromiso explícito de muchas revistas de corregir este problema. Incluso, un estudio reciente⁴⁹ muestra que incluso el tiempo desde que finaliza el reclutamiento de sujetos hasta la publicación también es menor para los estudios con resultados positivos.

HETEROGENEIDAD

En el modelo de efectos fijos habitualmente utilizado para estimar el efecto agrupado en el meta-análisis, la premisa es que todos los estudios son meras réplicas del mismo experimento, o lo que es lo mismo se asume que todos los ensayos son similares por el hecho de compartir el efecto del mismo tratamiento, por lo que las diferencias observadas se deben únicamente al azar y al error de muestreo. Por el contrario, al utilizar el modelo de efectos aleatorios se incorpora además la potencial heterogeneidad del efecto del tratamiento entre los diferentes estudios, al asumir que cada estudio estima un único efecto del tratamiento, aunque esta aproximación no explica las razones de dicha heterogeneidad. Aunque aún sigue en pie un arduo debate sobre cuál es el mejor método de análisis para estimar el efecto agrupado en el meta-análisis, los resultados suelen ser concordantes con ambos métodos a menos que exista una importante heterogeneidad entre los ensayos incluidos.

ANALISIS DE SENSIBILIDAD

El análisis de sensibilidad se podría definir como la valoración del impacto sobre las conclusiones del hecho de variar los métodos de análisis, como por ejemplo modificar criterios de inclusión, o bien las técnicas estadísticas, como aplicar diferentes test o realizar análisis ponderados por covariables consideradas de interés^{33,34,36}. Los factores determinantes de un análisis de sensibilidad pueden surgir del mismo proceso de realización de un meta-análisis, como por ejemplo las diferencias entre estudios que justifiquen la heterogeneidad de los mismos. En ocasiones, incluso el poner en evidencia la heterogeneidad y sobre todo el ser capaces de identificar las variables responsables de las diferencias observadas puede tener más importancia que la obtención de una medida global del efecto, por la información que pueden proporcionar sobre las diferencias de los efectos del tratamiento en diferentes estudios o subgrupos.

En este sentido, con el fin de explicar la heterogeneidad presente en muchos casos y para poder individualizar las recomendaciones derivadas de los meta-análisis, se han desarrollado en los últimos años técnicas estadísticas multivariable como la meta-regresión o el análisis de los datos individuales de cada paciente en cada ensayo⁵⁰ cuyas principales ventajas y limitaciones se resumen a continuación. La mega-regresión consiste en identificar las posibles relaciones existentes entre el efecto de

la intervención, que es la variable dependiente, con las covariables de interés denominadas variables independientes, para explorar la importancia de las diferencias entre los factores de riesgo en los diversos estudios incluidos en el meta-análisis. En este caso la unidad de observación no es el individuo, sino el estudio o el subgrupo previamente definido y la variable independiente puede ser constante entre los diferentes estudios o bien representar una medida que resuma los valores individuales de cada estudio. Aunque este enfoque proporcione mayor información al permitir describir los efectos multidimensionales de varios factores actuando conjuntamente y sugiera en ocasiones las posibles razones de la heterogeneidad, presenta varios problemas como son el que una asociación significativa no prueba causalidad, la posibilidad de incurrir en la denominada falacia ecológica cuando las variables de resumen de los datos no representan al paciente individual, y lo que es más importante, la ausencia frecuente de los datos que permitan definir las covariables al no estar disponibles en los ensayos originales e incluso la selección subjetiva de las mismas.

La manera de superar las limitaciones de la meta-regresión pasa por el análisis con datos individuales de los sujetos en cada estudio, lo que proporciona el máximo de información sobre las covariables potencialmente responsables de la heterogeneidad y aunque añade una gran complejidad, tiempo y esfuerzo al desarrollo del meta-análisis, sin embargo permite análisis pormenorizados de supervivencia como se ha mostrado recientemente para el cáncer de mama⁵¹. Normalmente permite escapar al sesgo de selección de las covariables, si estas están disponibles en todos los ensayos. También se ha utilizado la estrategia de utilizar como marcador, en lugar de alguna de las covariables que faltan, la tasa de eventos en el grupo control que es una medida global que refleja los múltiples factores como severidad del proceso, duración de seguimiento, efecto de tratamientos concomitantes que contribuyen a definir el riesgo de ocurrencia del evento en la población de estudio. Aunque esta estrategia es menos potente que la utilización de datos individuales, es una buena aproximación para describir las causas de heterogeneidad entre los diferentes ensayos.

CONCLUSIONES

Para concluir, desearía resaltar el papel complementario de los ensayos clínicos controlados y de los meta-análisis con el establecimiento de una escala creciente de complejidad y conocimiento. Así, mientras que los ensayos clínicos controlados de pe-

queño tamaño servirían para estudiar mecanismos fisiopatológicos, los meta-análisis de ensayos clínicos controlados de pequeño tamaño podrían generar hipótesis para futuros estudios. En el siguiente escalón, los ensayos clínicos controlados de gran tamaño servirían para obtener respuestas válidas de forma global y finalmente poder conseguir con los meta-análisis de ensayos clínicos controlados de gran tamaño, una estimación no sesgada y precisa del efecto del tratamiento además de poder explorar las potenciales interacciones y modificaciones del efecto de la intervención en los diferentes subgrupos. Finalmente, y volviendo a la idea con la que se iniciaban estas reflexiones de definir criterios objetivos de clasificación de los estudios en función de la valoración crítica de su metodología y de la interpretación de sus resultados, recomendar la lectura de un reciente artículo⁵² del grupo promotor de la medicina basada en la evidencia que insiste en algunos de los conceptos que se han esbozado en las líneas precedentes como son la utilización de normas metodológicas estrictas para la realización de los meta-análisis, la diferencia entre relevancia clínica y significación estadística, la precisión de la estimación del efecto y finalmente la diferencia del efecto de la intervención en diferentes subgrupos en función del nivel de riesgo, para establecer una graduación objetiva de las recomendaciones en la implementación de las intervenciones terapéuticas.

BIBLIOGRAFIA

- Hulley SB, Cummings SR: Diseño de la investigación clínica: un enfoque epidemiológico. Ed. Doyma, S.A., Madrid, 1993.
- Barrio V, Madero R: Estadística en la investigación en medicina. *Nefrología* 9: 236-242, 1989.
- Gardner MJ, Altman DG, ed: Statistics with confidence: confidence intervals and statistical guidelines. London, British Medical Journal, 1989.
- Medical Research Council: Streptomycin treatment of pulmonary tuberculosis. *BMJ* II: 769-782, 1948.
- Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG: Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 34: 585-612, 1976.
- Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG: Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *Br J Cancer* 35: 1-39, 1977.
- Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group: User's guides to the medical literature, II: how to use an article about therapy or prevention, A: are the results of the study valid? *JAMA* 270: 2598-2601, 1993.
- Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group: User's guides to the medical literature, II: how to use an article about therapy or prevention, B: what were the results and will they help me in caring for my patients? *JAMA* 271: 59-63, 1994.
- Rothman KJ, Michels KB: The continuing unethical use of placebo control. *N Engl J Med* 331: 394-398, 1994.
- Taubes G: Use of placebo controls in clinical trials disputed. *Science* 267: 25-26, 1995.
- Jones J, Jarvis P, Lewis JA, Ebbutt AF: Trials to show equivalence. *BMJ* 313: 36-39, 1996.
- Senn S: Are placebo run ins justified? *BMJ* 314: 1191-1193, 1997.
- Pocock SJ, Hughes MD, Lee RJ: Statistical problems in the reporting of clinical trials. *N Engl J Med* 317: 426-432, 1987.
- Pitt B, Segal R, Martínez FA, Meurers G, Cowlwy AJ, Thomas I, Deedwania PC, Ney DE, Snavely DB, Chang PI: Randomised trial of losartan versus captopril in patients over 65 with heart failure (Evaluation of Losartan in the Elderly study, ELITE). *Lancet* 349: 747-752, 1997.
- Yusuf S, Wittes J, Probstfield J, Tyroler HA: Analysis and interpretation of treatment effects in subgroups of patients in randomised clinical trials. *JAMA* 266: 93-98, 1991.
- Oxman AD, Guyatt GH: A consumer's guide to subgroup analysis. *Ann Intern Med* 116: 78-84, 1992.
- Schneider B: Analysis of clinical trial outcomes: alternative approaches to subgroup analysis. *Controlled Clin Trials* 10: 176S-186S, 1989.
- Dans AL, Dans LF, Guyatt GH, Richardson S, for the Evidence-Based Medicine Working Group: User's guides to the medical literature, XIV: how to decide on the applicability of clinical trial results to your patient. *JAMA* 279: 545-549, 1998.
- Rothwell PM: Can overall results of clinical trials be applied to all patients? *Lancet* 345: 1616-1619, 1995.
- Glasziou PP, Irwing LM: An evidence based approach to individualising treatment. *BMJ* 311: 1356-1359, 1995.
- Sharp SJ, Thompson SG, Altman DG: The relation between treatment benefit and underlying risk in meta-analysis. *BMJ* 33: 735-738, 1996.
- Frieman JA, Chalmers TC, Smith H Jr, Kuebler RR: The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial: survey of 71 negative trials. *N Engl J Med* 299: 690-694, 1978.
- DerSimonian R, Charette J, McPeck B, Mosteller F: Reporting on methods in clinical trials. *N Engl J Med* 306: 1332-1337, 1982.
- Begg C, Cho M, Eastwood S y cols.: Improving the quality of reporting of randomised controlled trials: the CONSORT statement. *JAMA* 276: 637-639, 1996.
- The European Agency for the Evaluation of Medical Products (EMA): Human Medicines Evaluation Unit. 1995/1996 Guidelines on the efficacy of medical products for human use issued by the CPMP/Efficacy Working Party. <http://www.eudra.org/emea.html>.
- Savulescu J, Chalmers I, Blunt J: Are research committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 313: 1390-1393, 1996.
- Henry D, Hill S: Comparing treatments: comparison should be against active treatments rather than placebos. *BMJ* 310: 1279, 1995.
- Chalmers I, Gray JAM, Sheldon T: Handling scientific fraud: prospective registration of health care research would help. *BMJ* 311: 262, 1995.
- Stern JM, Simes RJ: Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 315: 640-645, 1997.
- Light RJ, Smith PV: Accumulating evidence: procedures for resolving contradictions among different research studies. *Harvard Educ Rev* 41: 429-471, 1971.
- Glass GV: Primary, secondary, and meta-analysis of research. *Educ Res* 5: 3-8, 1976.

CONTROVERSIAS METODOLOGICAS SOBRE ENSAYOS CLINICOS

32. Cochran WG: Planning and analysis of observational studies. John Wiley & Sons, New York, 1983.
33. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC: Meta-analysis of randomized controlled trials. *N Engl J Med* 316: 450-455, 1987.
34. L'Abbé KA, Detsky AS and O'Rourke K: Meta-analysis in clinical research. *Ann Int Med* 107: 224-233, 1987.
35. Boissel JP, Blanchard J, Panak E, Peyrieux JC, Sacks H: Considerations for the meta-analysis of randomized clinical trials. Summary of a panel discussion. *Controlled Clin Trials* 10: 254-281, 1989.
36. Jenicek M: Meta-analysis in medicine. Where we are and where we want to go. *J Clin Epidemiol* 42: 35-44, 1989.
37. Chalmers I, Altman DG, eds: Systematic reviews. London: BMJ Publishing Group, 1995.
38. Oxman AD, Cook DJ, Guyatt GH, for the Evidence-Based Medicine Working Group: User's guides to the medical literature, VI: how to use an overview. *JAMA* 272: 1367-1371, 1994.
39. Pogue J, Yusuf S: Overcoming the limitations of current meta-analysis of randomised controlled trails. *Lancet* 351: 47-52, 1998.
40. Cook DJ, Sackett DL, Spitzer WO: Methodological guidelines for systematic reviews of randomised controlled trials in health care from the Postdam Consultation in Meta-analysis. *J Clin Epidemiol* 48: 167-171, 1995.
41. Villar J, Carroli G, Belizan Jm: Predictive ability of meta-analysis of randomised controlled trials. *Lancet* 345: 772-776, 1995.
42. Cappelleri JC, Ionnadis JPA, Schmid CH y cols.: Large trials vs meta-analysis of small trials: how do they results compare? *JAMA* 276: 1332-1338, 1996.
43. LeLorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F: Discrepancies between meta-analysis and subsequent large, randomized, controlled trials. *New Engl J Med* 337: 536-542, 1997.
44. Egger M, Smith GD, Schneider M, Minder C: Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315: 629-634, 1997.
45. CLASP (Collaborative Low-dose Aspirin Study in Pregnancy) Collaborative Goup. CLASP: a randomised trial of low dose aspirin for the prevention and treatment of pre-eclampsia among 9364 pregnant women. *Lancet* 343: 619-629, 1994.
46. ECPPA (Estudo Collaborativo para Prevenção da Pré-ecamp-sia com Aspirina) Collaborative Group. ECPPA: randomised trial of low dose aspirin for the preventeion of maternal and fetal complications in high risk pregnant women. *Br J Obstet Gynecol* 103: 39-47, 1996.
47. Caritis S, Sibai B, Hauth J, Lindheimer MD y cols.: Low-dose aspirin to prevent preeclampsia in women at high risk. *N Engl J Med* 38: 701-705, 1998.
48. Imperiale TF, Petrusis AS: A meta-analysis of low-dose aspirin for the prevention of pregnancy induced hypertensive disease. *JAMA* 266: 260-264, 1991.
49. Ionnadis JPA: Effect of the statistical significance of results on the time to completion and publication of randomised efficacy trials. *JAMA* 279: 281-286, 1998.
50. Lau J, Ionnadis JPA, Schmid CH: Summing up evidence: one answer is not always enough. *Lancet* 351: 123-127, 1998.
51. Early Breast Cancer Trialists' Collaborative Group: Effects of radiotherapy and surgery in early breast cancer: an overview of the randomized trails. *N Engl J Med* 333: 1444-1455, 1995.
52. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ, for the Evidence-Based Medicine Working Group: User's guides to the medical literature, IX: a method for grading health care recomendations. *JAMA* 274: 1800-1804, 1995.