



Revisiones sistemáticas y metaanálisis. Elementos de la investigación integrativa

Héctor A. Baptista-González*

* Hematología Perinatal. Subdirección de Investigación Clínica. Instituto Nacional de Perinatología Isidro Espinosa de los Reyes.

RESUMEN

Con este artículo se inicia una serie de documentos y reflexiones acerca de las herramientas de revisión sistemática (RS) y metaanálisis (MA) aplicables a la evaluación de las intervenciones clínicas. La implementación exitosa requiere del entendimiento de la información que proporcionan las RS y MA, que el nivel jerárquico que toma decisiones en salud (haciendo a un lado la resistencia al cambio) tenga en cuenta esta modificación en el paradigma y genere los espacios adecuados para su desarrollo y aplicación. Al finalizar esta serie de artículos, es nuestra intención que los lectores, mediante la valoración crítica de las fuentes de información secundaria, en respuesta a preguntas clínicas concretas, se avituallen de los elementos necesarios para el empleo juicioso, razonado y crítico de la estructura que debiera de regir en buena parte de las intervenciones de nuestro quehacer cotidiano: la Guía de Práctica Clínica.

Palabras clave: Medicina basada en la evidencia, medicina basada en la práctica, metaanálisis, aprendizaje basado en problemas

ABSTRACT

This is the first article of a sequence of documents and reflections about systematic review (SR) and meta-analysis (MA) tools available for the evaluation of clinical interventions. The implementation success needs the SR and MA information understanding, their influence over the decision making at a hierarchic level and the fabrication of the necessary spaces for their development and application. At the end of the sequence, the reader will find answer to concret clinical questions by a critical evaluation of the secondary information tools and to develop the necessary elements for the judgments, reasonable and critical use of clinical practice guides.

Key words: Evidence based medicine, evidence based-practice, meta-analysis, problem-based learning.

Las actividades de la práctica clínica cotidiana y de la enseñanza en medicina emplean una plétora de técnicas didácticas, donde se incluyen métodos basados en lectura y habilidades, práctica conceptualizada y otros más basados en la tecnología o bien en la investigación (casos clínicos).

De primera instancia, tales métodos parecen complementarse unos a otros. Sin embargo, la manera en que aprende la gente es a partir de preguntas como: “¿cuál es la mejor técnica para actualizar el conocimiento?, ¿cuál es la mejor técnica para aprender?, ¿cuál es la mejor técnica para enseñar?” El aprendizaje basado en problemas (ABP) se fundamenta en la teoría cognoscitiva y tiene sus orígenes en la educación médica; es un acercamiento útil para

los estudiantes de medicina sobre cómo pensar críticamente y solucionar problemas cotidianos.²

La medicina basada en evidencias científicas (MBE) es un método para practicar y enseñar la medicina clínica; es el empleo consciente, explícito y juicioso de la mejor evidencia actual en la toma de decisiones sobre el cuidado sanitario de los pacientes.¹ La MBE emplea dos herramientas poderosas para valorar críticamente la utilidad, eficacia y seguridad de las intervenciones etiológicas, diagnósticas, terapéuticas y pronósticas, que día a día ocurren en la cabecera del paciente, en el consultorio o en la toma de decisiones de políticas de salud.

Con el presente, se inicia una serie de artículos que tratan los temas de RS y MA divididos en capí-

tulos. En el capítulo inicial, se discuten las definiciones operacionales de ambos conceptos y se acompaña al final de un glosario de los términos más empleados. Es necesario tener los elementos básicos para la interpretación de los resultados de las RS y MA, sin tener que recorrer los complicados caminos de la estadística; para lograr lo anterior, se decidió agregar un capítulo segundo, sobre la interpretación de los estudios de RS y MA para inexpertos. El tercer tema es sobre desarrollo técnico y metodológico de las RS y MA, que abarca mayor profundidad en sus conceptos y las estrategias más comunes para su abordaje e interpretación. El cuarto capítulo describe un análisis sobre la mecánica de pensamiento que se requiere para elaborar una adecuada pregunta clínica, que ayude a definir el problema particular que tenemos en el consultorio o en la cabecera del paciente. La pregunta clínica es el recurso que cotidianamente se emplea para resolver el problema particular ante cada paciente. Se denomina pregunta clínica estructurada a aquella que tiene una serie de componentes lógicos que permiten organizarla y plantearla en términos que pueden ser respondidos con la utilidad deseada para el paciente. El plantear la pregunta correcta, buscar la información con la estrategia adecuada, evaluar los datos obtenidos con la metodología aplicable y obtener una respuesta útil, son los objetivos de esta serie de artículos.⁴ El quinto y sexto capítulo tratan los temas del empleo de los recursos de Internet, para la búsqueda de bases de datos electrónicas y el empleo de programas computacionales diseñados para la elaboración de RS y MA. Esta serie concluye con el acercamiento de los lectores a las herramientas de la valoración crítica de la calidad de las RS y MA, publicadas en áreas específicas de la medicina.

De ninguna manera se pretende agotar los temas de RS y MA. Por el contrario, se intenta explicar de una manera accesible y sencilla los elementos críticos de estas herramientas. Se parte del hecho de que la profundidad en el conocimiento y el nivel de experiencia es altamente variable. Deberá quedar claro para el lector que llevar una intervención clínica, producto de una RS o MA, es materia que se discutirá en otros artículos. Aún se tiene que luchar contra la resistencia al cambio en el conocimiento y la práctica, lo que ocurre en toda la jerarquía médica, desde las más altas autoridades (donde se establecen las prioridades en las políticas sanitarias), pasando

por los directivos de hospitales, jefes de servicio, médicos adscritos y residentes.

El lector deberá estar atento a identificar, desde el inicio, que el empleo del método científico es el criterio general que se aplica. Ni las creencias, preferencias, gustos personales o glorias vanas en la toma de decisiones clínicas son temas que deban tratarse en esta revista, ni mucho menos por el autor. Por el contrario, se busca estimular el sentido crítico en la evaluación de la información, pero también es necesario desarrollar las estrategias de pensamiento y razonamiento en el quehacer médico cotidiano. Sin duda alguna, las herramientas empleables para tal fin son abundantes. Una de ellas, a la que se recurrirá con cierta frecuencia, es la heurística, definida como la capacidad de un sistema para realizar de forma inmediata innovaciones positivas para sus fines. La capacidad heurística es el arte y la ciencia de resolver problemas mediante la creatividad y el pensamiento lateral o divergente. Como disciplina científica, la heurística es aplicable al razonamiento clínico e incluye la elaboración de medios, principios, reglas, estrategias y programas que faciliten la búsqueda de vías de solución a problemas; es decir, resolver tareas de cualquier tipo para las que no se cuente con un procedimiento algorítmico de solución. Los procedimientos heurísticos son formas de trabajo y pensamiento que apoyan la realización consciente de actividades mentales exigentes. El método o procedimiento heurístico se divide en principios, reglas y estrategias. Los principios heurísticos constituyen sugerencias para encontrar directamente la idea de solución; posibilita determinar los medios y la vía de solución. Dentro de estos principios, se destacan la analogía y la reducción. Las reglas heurísticas actúan como impulsos generales dentro del proceso de búsqueda y ayudan a encontrar los medios para resolver los problemas. Se incluyen acciones como: separar lo dado de lo buscado, confeccionar figuras de análisis: esquemas, tablas, mapas, etc., representar magnitudes dadas y buscadas con variables, determinar si se tienen fórmulas adecuadas, utilizar números (estructuras más simples) en lugar de datos y reformular el problema. Las estrategias heurísticas se comportan como recursos organizativos del proceso de resolución, que contribuyen especialmente a determinar la vía de solución del problema abordado. Existen dos estrate-

gias: 1. El trabajo hacia adelante: Se parte de lo dado para realizar las reflexiones que han de conducir a la solución del problema; 2. El trabajo hacia atrás: Se examina primeramente lo que se busca y, apoyándose en los conocimientos que se tienen, se analizan posibles resultados intermedios de lo que se puede deducir lo buscado, hasta llegar a lo dado.

Cuando el Maestro Samuel Dorantes realizaba su pase de visita en la Sección de Hematología del antiguo Hospital Infantil de México, procedía a la revisión clínica minuciosa de cada uno de sus pacientes y, posteriormente, en la sala de juntas, se efectuaba el análisis de los problemas anteriores y nuevos que sus pacientes mostraron, para posteriormente discutir en sesiones kilométricas los hallazgos encontrados. Al paso del tiempo —y espero tener oportunidad de confirmarlo directamente con el maestro Dorantes—, entiendo que empleaba los principios del pensamiento heurístico en su trabajo clínico, mismo que el matemático húngaro George Pólya sintetizó magistralmente:⁵ “*Si no consigues entender un problema, dibuja un esquema (por supuesto que se incluye el proceso mental). Si no encuentras la solución, haz como si ya la tuvieras y mira qué puedes deducir de ella (razonando hacia atrás). Si el problema es abstracto, prueba a examinar un ejemplo concreto. Intenta abordar primero un problema más general (paradoja del inventor: el propósito más ambicioso es el que tiene más posibilidades de éxito)*”.

Resulta muy difícil, desde la perspectiva del autor, lograr un avance en nuestras actividades médicas, siempre en beneficio y seguridad de nuestros pacientes, sin tener que lograr el cambio en el paradigma, cambio en las actitudes hacia nuestro quehacer diario.

Con estos antecedentes, el objetivo de la presente serie es establecer las bases para la interpretación adecuada de las intervenciones clínicas basadas en la evidencia científica disponible, mediante el empleo de las herramientas de revisión sistemática y metaanálisis.

Es frecuente la contradicción de conjuntar una buena práctica clínica diaria con las actividades de enseñanza. El agudo aforismo “*quien deja la clínica, se pone a enseñar o administra*”, exhibe esta incongruencia. Así que, careciendo de una preparación académica formal para participar en la enseñanza de la medicina, que por excelencia debería ser tutelar, queda como un reto, debido a las limitacio-

nes del autor. Así pues, el camino es difícil, la intención queda clara, los resultados estarán a su juicio.

INTRODUCCIÓN

El esquema de evaluación sistemática y estructurada del conocimiento científico en la medicina, conduce a lo que llamamos comúnmente evidencia clínica. La generación de elementos que permiten la construcción del conocimiento o evidencia clínica tiene sus principios elementales en la investigación básica. Esta etapa se estructura de manera escalonada a partir de la formulación de un concepto, definición, idea o hipótesis, producto directo en la continuidad de la investigación o bien como resultado de un hallazgo casual, no esperado, no buscado, conocido con serendipia. La investigación básica, instrumentando el método científico, continúa con experimentación en animales, cuyos resultados enriquecen o debilitan las hipótesis previamente propuestas. Si en el proceso de investigación se obtienen resultados positivos o favorables en la etapa básica, entonces se justifica (pero nunca antes) el estudio en seres humanos (investigación clínica), que avanza en el reporte de series de casos, hasta la aparición del diseño metodológico experimental en los casos y controles, cohortes o estudios clínicos controlados de mayor complejidad y rigor metodológico. Finalmente, una vez que se concluye el estudio en pacientes, se puede avanzar en el conocimiento al integrar los resultados de los estudios, tratados bajo diseños metodológicos y estadísticos peculiares. Así se pasa a la etapa de la investigación integrativa, donde no se analiza directamente a los pacientes; más bien se trata del manejo de la información de los resultados obtenidos en los diferentes estudios clínicos. Al final de esta pirámide, se logra el objetivo de alcanzar la síntesis conjuntada en recomendaciones desarrolladas de forma sistemática para ayudar a profesionales y pacientes a tomar decisiones informadas sobre la atención sanitaria más apropiada, seleccionando las opciones diagnósticas y/o terapéuticas más adecuadas o sea las Guías de Práctica Clínica (*Figura 1*).

El sentido de la construcción del conocimiento científico y la búsqueda de la evidencia clínica se relacionan con tres etapas en la investigación: básica, clínica e integrativa. En la investigación clínica,



Figura 1. Pirámide de la evidencia clínica.

la materia central es el estudio en la evaluación de las intervenciones en los seres humanos; en la investigación integrativa, el elemento central lo constituye información obtenida en la calidad y diseño metodológico de los estudios evaluados.

La información se obtiene de las fuentes primarias, secundarias y terciarias. Las fuentes primarias presentan de primera intención el estado de la investigación científica y sus resultados. Las fuentes de información secundaria, describen los resultados generados en las fuentes primarias y este conocimiento requiere ser periódicamente organizado y evaluado por la comunidad clínica y científica para otorgarle un sentido de estado de la ciencia. Las fuentes de información terciaria incluyen a los libros de texto como las fuentes habituales de información.

La revisión sistemática (RS) y metaanálisis (MA) pertenecen a la información secundaria y son conceptos relativamente recientes. En contraposición a la revisión narrativa, la revisión sistemática viene a dotar de un mayor rigor tanto al proceso de selección de los trabajos como a la posterior integración y análisis de los resultados.⁶ La revisión sistemática es la síntesis de los resultados de varios estudios primarios mediante técnicas que limitan los sesgos y el error aleatorio. Estas técnicas incluyen la búsqueda de todos los estudios potencialmente relevantes y el uso de criterios explícitos y fiables en la selección de las investigaciones, la evaluación crítica, sintetizar la evidencia, evaluando el efecto individual de cada estudio utilizando unos métodos estadísticos que den más peso en el cálculo del efecto combinado, a aquellos estudios con tamaño muestral mayor y así obtener conclusiones sobre el efecto

de una intervención específica (etiología, diagnóstica, terapéutica o pronóstica). Hay diferentes tipos de ponderaciones: inverso de la varianza, por la distribución de los datos (Mantel-Haenszel), por la varianza de los casos esperados en la referencia.⁷

La RS muestra las ventajas de proporcionar la síntesis de información respecto de una pregunta clínica específica, pues resume en un solo valor numérico toda la evidencia relacionada a un tema, aumentando la potencia estadística y la precisión del estimador puntual; al combinar la información de diversos estudios, permiten analizar la consistencia de los resultados. Muchos de los estudios tomados individualmente son demasiado pequeños incluso para detectar efectos modestos pero importantes, es decir, tienen un poder estadístico insuficiente.⁸ Combinando los estudios que han intentado responder a la misma pregunta se consigue aumentar el tamaño muestral y así su poder estadístico.⁹ Otra ventaja importante es que un efecto similar en diferentes ámbitos y utilizando diseños distintos (como los criterios de inclusión y exclusión diferentes para los pacientes que entran en cada estudio) puede darnos una idea de cuan robustos y aplicables son los resultados de nuestra RS a otros ámbitos. Si los resultados de los estudios individuales varían mucho entre uno y otro, se debe identificar cuál es la fuente que produce las diferencias en los resultados entre los estudios. Adicionalmente, genera el planteamiento de nuevas hipótesis para futuros estudios, junto con la detección de áreas en que la evidencia científica es escasa, cuantifica la magnitud del sesgo de publicación y resulta con menos costos que otro tipo de estudio y con un mayor nivel de impacto.

La evaluación crítica de los estudios de RS y MA para guiar la práctica clínica basada en la evidencia tiene diferentes maneras de ser aplicada. Los pasos más simples involucran:⁸

- a) Establecer la validez de los resultados de la revisión.
- b) Los resultados en sí mismos.
- c) La posibilidad de que esos resultados puedan ser aplicables al cuidado cotidiano del paciente.

Dentro de las herramientas utilizadas en el proceso de revisión, el metaanálisis constituye una de las piezas fundamentales y define los procedimientos cuantitativos de análisis de los datos extraídos de los trabajos de investigación revisados, con el fin de obtener conclusiones estadísticas a partir de ellos. El MA cumple el doble objetivo de estimar el valor medio global de la cantidad medida (por ejemplo eficacia del tratamiento) y analizar la heterogeneidad de ésta entre los diferentes estudios. Ésta puede proceder de muchas fuentes (más numerosas en los estudios de observación que en los diseños experimentales): características de la población del estudio (por ejemplo, el riesgo subyacente del efecto o subgrupos diferentes de alto o bajo riesgo), variaciones en el diseño del estudio (tipo de diseño, métodos de selección, fuentes de información, manera de reunir la información), diferentes métodos estadísticos y distintos esquemas de ajuste de los factores de confusión (*Figura 2*). Si hay heterogeneidad, el estimador ponderado no tiene sentido, ya que ello significa que hay más de una magnitud de efecto verdadera en los estudios que se combinan.⁹

LIMITACIONES DEL METAANÁLISIS Y LAS REVISIONES SISTEMÁTICAS (RS Y MA)

Estas fases de la investigación integrativa presentan limitaciones metodológicas inherentes a este tipo de estudios.¹⁰ Las RS muestran un grado de complejidad considerable, pues se debe considerar que los estudios retrospectivos están sujetos a sesgo en las diversas etapas del proceso. Por otro lado, está limitada por la cantidad y calidad de los estudios previamente realizados, la heterogeneidad de los mismos y es un proceso largo y engorroso, requiere tiempo y dedicación.

Las limitaciones metodológicas más relevantes que afectan los resultados e interpretación de los estudios de RS y MA, son:

1. Sesgo de publicación.⁹ Viene derivado del hecho de que muchos trabajos científicos, en su mayoría con resultados negativos (aquellos que no hallan diferencias significativas o con resultados en contra de la hipótesis de estudio o de lo habitualmente establecido) no llegan a publicarse. Todo ello condiciona los resultados de una búsqueda bibliográfica y puede dar lugar a resultados sesgados en un metaanálisis. Las causas de este sesgo pueden ser por una deficiente estrategia y rigor metodológico de los autores, cuando se cede ante las presiones institucionales o económicas, o en contratos industriales para obtener resultados favorables al producto, equipo o dispositivo en evaluación. Adicionalmente, los directores de revistas médicas tienen más tendencia a rechazar su publicación, porque los resultados negativos muestran menor impacto. También se ha comprobado que los estudios clínicos en los que no se hallan diferencias, tardan más tiempo en ser publicados.^{11,12} Cuando un mismo estudio ha sido publicado más de una vez, en formas aparentemente distintas, la publicación duplicada también puede introducir un sesgo, que favorece la tendencia del resultado duplicado. Adicionalmente, hay diferencias en la calidad de informes de los metaanálisis publicados en la base de datos de Cochrane de revisiones sistemáticas que es superior a los MA reportados en revistas periódicas.¹³

La determinación del riesgo de sesgo en estudios individuales en una revisión sistemática, se puede hacer usando componentes individuales o resumiendo el estudio de calidad en una cuenta total. En un reporte que examinó las metodologías empleadas por los 50 grupos de trabajo que elaboran las RS de la Colaboración Cochrane, centraron sus evaluaciones de los artículos basados en dos estrategias distintas: evaluar la calidad de los componentes de cada artículo o bien emplear una escala de calificación de la calidad metodológica de los estudios. En 41 de los grupos de revisión (82%), se recomendó la evaluación de la calidad utilizando los componentes del documento y nueve usando una escala. Todos los grupos que recomendaban componentes sugirieron determinar metodología de la asignación, sólo comparada con dos de los grupos de escalas de recomendación. En 21 grupos (42%) se recomendó análisis

de la intención de tratar. Solamente 28 grupos (56%) tenían recomendaciones específicas para usar la evaluación analítica de los componentes del artículo, con análisis de sensibilidad, calidad como umbral de la inclusión y análisis de subgrupos, como el método más recomendado. Las escalas recomendadas tenían problemas en los artículos individuales. Si bien las recomendaciones de algunos grupos no fueron basadas en evidencia empírica y muchos grupos no tenían ninguna recomendación en cómo utilizar la evaluación de calidad de los componentes del artículo. Es necesario que todos los grupos de revisión de la Cochrane, se refieran al Manual para las Revisiones Sistemáticas de intervenciones, como un marco de referencia general en la elaboración de revisiones sistemáticas.¹⁴

Tal como ocurre en la investigación clínica o en la básica, la elaboración de un documento de investigación integrativa, no está exento de sesgo. Al evaluar la evidencia referente al sesgo y confusión presentes en la conducción de las revisiones sistemáticas. Una RS reciente¹⁵ valora el riesgo del sesgo en la conducción de RS, empleando el índice de Oxman y de Guyatt. Se identificaron diez RS; todas presentaron sesgo relacionado con la búsqueda de la evi-

dencia (sesgo de publicación). Un estudio señaló la presencia de sesgo relacionando la obtención de datos de los estudios incluidos (sesgo en el reporte de resultados). Para minimizar el sesgo, los datos sugieren incluir material no publicado, la búsqueda manual de material adicional, evaluar múltiples bases de datos, evitar finalmente el sesgo de publicación, así como actualizar constantemente las RS.

2. Sesgo del idioma. Es un hecho persistente que en una proporción de metaanálisis, se ignoran los estudios clínicos publicados en idiomas distintos del inglés y se ha visto que la proporción de estudios con resultados es mayor en lenguas no inglesas.

3. Sesgo de selección. Es una de las principales críticas del MA. Se requiere definir con claridad los criterios de inclusión y exclusión de los estudios en la revisión y que éstos sean lo más objetivos posible. El propio autor del MA puede sesgar los resultados, ya que su criterio para incluir o excluir un estudio del análisis puede venir influenciado por los resultados del mismo. Otros aspectos que pueden comprometer la validez de los resultados de un MA son la calidad de los estudios originales incluidos, la variabilidad entre estudios o los errores en la fase de análisis. Las conclusiones del MA dependerán en

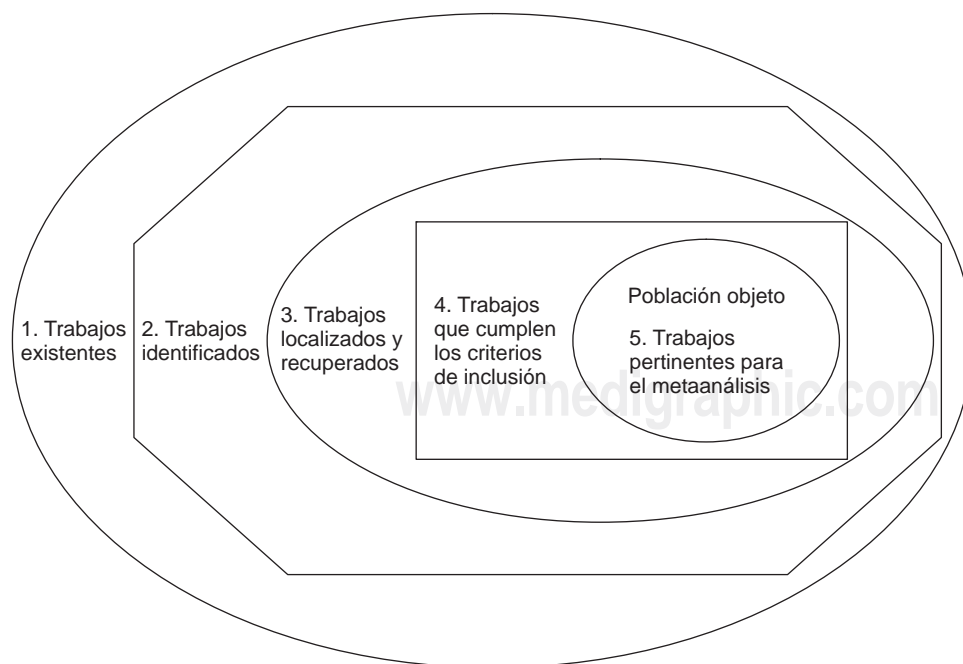


Figura 2. Identificación del material publicado para el desarrollo de un estudio de metaanálisis.

gran medida de la calidad de los estudios originales, de modo que al combinar resultados de investigaciones sesgadas o metodológicamente deficientes, también se incurrirá en un sesgo, a la hora de obtener un estimador global del efecto.

4. Heterogeneidad. Esto ocurre cuando los diferentes estudios que se combinan pueden afectar de una manera muy importante los resultados del MA. Suelen ser investigaciones realizadas en distintos contextos, con pacientes de características no necesariamente similares o incluso con resultados muy diferentes, lo que implica que no siempre será aconsejable realizar un MA. Deberán investigarse las posibles fuentes de heterogeneidad, su influencia en los resultados y la posibilidad de realizar un análisis por subgrupos.

En teoría, la agregación de datos de múltiples estudios clínicos debe potenciar la precisión y la exactitud de cualquier resultado agregado. No obstante, combinar datos implica presuponer que las diferencias entre estudios se deben al azar, cuando en realidad se deben a otros factores, como diferencias más o menos sutiles en las poblaciones participantes, en la definición y la medición de las variables, en el diseño y su calidad metodológica general. El MA puede generar resultados equívocos cuando ignore una heterogeneidad real entre los estudios, y, de manera inadvertida, borre los sesgos de uno, simplemente compensándolos con los de sentido contrario de otro.¹⁶

5. Diferencias en las medidas o técnicas de análisis estadístico. Con el empleo de técnicas de análisis erróneas o una definición incorrecta de las medidas que se desean combinar, los valores de significación, que no informan del sentido ni de la magnitud de la asociación, o los estadísticos de contraste, que tienen en cuenta el sentido del efecto, pero no de su magnitud, no deberían considerarse como medidas de entrada al MA.

A pesar del avance en el perfeccionamiento metodológico, algunos aspectos son objeto todavía de controversia: el primer objetivo debería ser la estimación de un resumen global de efecto promedio entre estudios o la identificación y estimación de las diferencias entre los efectos específicos de los mismos. El mayor problema radica en que un MA puede ofrecer una impresión falsa sobre la consistencia a través de los resultados ofrecidos de forma individual. Un problema frecuente es cómo actuar cuando los estudios que se intentan combinar son heterogéneos,

pues ofrecen resultados discordantes. Algunas de las opciones posibles son presentar los resultados individuales sin combinarlos, combinar los estudios utilizando técnicas estadísticas que incorporan su variabilidad (modelos de efectos aleatorios), o intentar evaluar cuáles son las causas de la heterogeneidad. Finalmente, se ha de insistir en que la calidad de los estudios de metaanálisis está limitada por la calidad de los estudios primarios. Las conclusiones obtenidas a partir de estudios sesgados estarán sesgadas, por muy alto que sea el número disponible.

6. Diferencias subjetivas en los autores de las RS y MA. Las discrepancias entre las conclusiones de diversos metaanálisis, se atribuyen a menudo a las diferencias metodológicas. Con cierta frecuencia se requiere evaluar la discordancia en los resultados del metaanálisis, cuando emplean la misma fuente primaria. En un reporte reciente,¹⁷ se envió a un grupo de expertos un paquete de artículos relacionados con la eficacia del magnesio endovenoso administrado en el periodo inmediato posterior al infarto al miocardio, para que luego de su revisión se contestara si la información era suficiente para demostrar el beneficio o no del magnesio. Los resultados fueron interesantes, pues existió desacuerdo considerable entre los expertos, para cada bloque de artículos y de la utilidad del fármaco. Estas discrepancias aumentaron proporcionalmente a la heterogeneidad de los resultados. Es decir, que la interpretación de los resultados de revisiones sistemáticas con metaanálisis incluye un componente subjetivo, que pueda llevar a conclusiones discordes, independientes de la metodología usada para obtener o para analizar los datos.

Las revisiones sistemáticas son las más provechosas si se actualizan periódicamente. Sin embargo, la información proveniente de fuentes de información primaria se agrega día a día, contrastando sus resultados respecto a los señalados en las RS. La incorporación de las intervenciones clínicas, derivadas de la evidencia publicada en RS o MA, a la práctica cotidiana, también tarda 2-4 años en ocurrir. Con cierta frecuencia, las conclusiones de la RS están ya superadas por la evidencia actual. Es importante señalar que no existe suficiente información sobre la uniformidad de criterios, que permitan establecer cuándo o cómo poner al día revisiones sistemáticas.¹⁸

GLOSARIO

(<http://www.medtrad.org/panacea/IndiceGeneral/n8-DelgadoRodriguez.pdf> http://www.guiasalud.es/docs/GLOSARIO_GS_feb_06.PDF)

Análisis costo-efectividad (Cost-effectiveness analysis). Forma de análisis socioeconómico mediante el cual se comparan intervenciones alternativas y en el que los costos se miden en términos económicos y los resultados se expresan en unidades no monetarias. Se describen los costos para alguna ganancia o pérdida en salud.

Análisis de sensibilidad (Sensitivity analysis). Existen diferentes tipos. En la valoración de la heterogeneidad, el análisis de sensibilidad trata de la inclusión/exclusión de estudios específicos. En el uso de técnicas estadísticas de combinación, consiste en la repetición del análisis de combinación mediante varios métodos, para valorar si se alcanzan los mismos resultados y si éstos se alteran por la calidad de los estudios individuales y el sesgo de publicación.

Análisis de supervivencia (Survival analysis). Método que permite determinar la probabilidad de un suceso (muerte, curación, recidiva, etc.) en diferentes intervalos de tiempo.

Análisis por subgrupos (Subgroup analysis). Método en el cual se analizan los datos de subgrupos de pacientes. Los subgrupos deben estar definidos previamente al inicio del estudio y normalmente guardan relación con factores pronósticos (sexo, edad, gravedad de la enfermedad, etc.).

Análisis multivariado. Es un tipo de análisis estadístico que trata de evaluar la asociación independiente de una variable con un evento, teniendo también en cuenta la participación simultánea de otras variables.

Calidad de un estudio (Study quality). Valoración global de un estudio según un protocolo probado y validado. Dado que las diferentes escalas de calidad dan resultados divergentes, deben interpretarse con precaución los análisis basados en una puntuación global de la calidad. Resulta más fácil en los ensayos clínicos que en los estudios de observación.

Diferencia de riesgos/Reducción de riesgo absoluto, RRA (Absolut risk difference). Es la diferencia absoluta entre los riesgos (incidencias)

entre dos grupos de comparación; un riesgo cero indica que no hay diferencia entre los grupos que se comparan. Para resultados indeseables, una diferencia menor de cero indica que la intervención fue eficaz al reducir el riesgo de aquel suceso.

Confusor. La presencia de un efecto confusor se observa cuando la asociación entre una exposición y el evento en estudio se modifica en función de la presencia o ausencia de un tercer factor. En estas circunstancias, la relación observada entre la exposición y el evento puede ser explicada parcial o totalmente por la variable confundidora. Para ser denominada como tal, este tercer factor debe asociarse independientemente, tanto al factor de exposición como al evento en estudio.

Efectividad (Effectiveness). Es la magnitud en la que una intervención (tratamiento, procedimiento o servicio) mejora los resultados para los pacientes en la práctica.

Efecto placebo (Placebo effect). Efecto en los resultados (mejoría o empeoramiento) que ocurre debido a las expectativas que tienen, el paciente y el investigador, de que la intervención producirá un efecto determinado.

Eficacia (Efficacy). Es el grado o la magnitud en la que una intervención (tratamiento, procedimiento o servicio) mejora los resultados para los pacientes en condiciones ideales (típicamente un ensayo clínico controlado con asignación aleatoria).

Eficiencia (Efficiency). Relación entre los resultados obtenidos y el costo de los recursos consumidos para obtener los resultados.

Error aleatorio (Sampling error). Es el error debido al azar. Los intervalos de confianza y los valores de p representan la posibilidad de los errores aleatorios, pero no de los errores sistemáticos (sesgos).

Error alfa (Alpha level). Es la probabilidad de un error tipo I que consiste en rechazar la hipótesis nula siendo verdadera. En la prueba de hipótesis, el nivel alfa es el umbral para definir la significación estadística. Por ejemplo, situando alfa en el nivel de 0.05 implica que los investigadores aceptan que hay un 5% de probabilidad de concluir incorrectamente que una intervención es efectiva cuando en realidad no tiene ningún efecto. El nivel alfa se establece normalmente en 0.01 ó 0.05.

Error beta (Beta level). Es la probabilidad de un error tipo II, que consiste en aceptar la hipótesis

nula siendo falsa. En la prueba de hipótesis, el nivel beta es la probabilidad de concluir incorrectamente que una intervención no es efectiva cuando sí lo es. Se considera que (1-beta) es la potencia de la prueba para detectar el efecto de una intervención si éste verdaderamente existe.

Estadístico Q (Q statistic). Es la prueba de la Chi cuadrada para valorar la heterogeneidad de los estudios incluidos en un MA, en la que la magnitud del efecto de cada estudio individual se compara con el estimador combinado. Por su validez, potencia estadística y facilidad de cálculo, esta prueba de heterogeneidad es la mejor elección.

Estimador (Estimator). Es la medida resumen calculada en una muestra. Por ejemplo: media, riesgo, riesgo relativo, etc. Los estimadores se usan para hacer inferencias sobre la población. Deberían ir acompañados de su correspondiente error estándar.

Estimación ajustada (Adjusted estimate). En un estudio, se denomina estimación ajustada a la realizada teniendo en cuenta los factores de confusión que pueden influir en los resultados. Los habituales son: estratificación, estandarización y modelos de regresión. La estimación ajustada es lo opuesto a estimación cruda.

Estimación cruda (Crude estimate). Es la estimación obtenida sin tener en cuenta factores de confusión.

Estrategia de búsqueda (Search strategy). Es la traducción de una pregunta clínica al lenguaje documental, identificando los términos específicos y combinándolos adecuadamente.

Estratificación (Stratification). Es el método de control de los factores de confusión, en el cual se presentan los resultados de acuerdo con subgrupos (estratos de pacientes). Procedimiento para calcular estimaciones o realizar contrastes de hipótesis, para cada nivel, o estrato, de una variable categórica y después calcular una estimación global para todos los estratos. El método más usado es el de Mantel-Haenszel.

Factor de confusión (Confounding factor). Es la variable que está simultánea e independientemente asociada al factor de riesgo y a la enfermedad estudiada y puede modificar el resultado del estudio. Ej: se está investigando a un grupo de personas con sobrepeso y a un grupo de personas sin sobrepeso. Los dos grupos tienen diferentes edades, por lo que una diferencia en el riesgo de enfermedad car-

diaca podría no ser debida al sobrepeso, sino a la propia edad. La edad puede actuar como factor de confusión. Las estimaciones crudas no son válidas en estas circunstancias.

Factor de riesgo (Risk factor). Una característica o estilo de vida de una persona, o de su entorno, que incrementa la probabilidad de ocurrencia de una enfermedad.

Gráficos en árbol de navidad (Christmas tree plots). Gráfico que se utiliza para detectar el sesgo de publicación, donde el eje de ordenadas representa el tamaño del estudio (o el error estándar del parámetro que se pondera en sentido decreciente) y en las abscisas el parámetro que mide la magnitud del efecto. Se traza un eje vertical que pasa por el valor ponderado global. La asimetría con relación al eje (sobre todo en la parte baja del árbol) indica sesgo de publicación.

Gráficos en embudo (Funnel plots). Es el conjunto de métodos gráficos para representar la existencia de un posible sesgo de publicación. Muestra la relación entre la magnitud del efecto (eje de ordenadas) y el tamaño del estudio (eje de abscisas), que se puede medir de distintas maneras (error estándar de la magnitud del efecto, su inverso, el tamaño de la muestra o el número de efectos observados). Se dibuja una línea horizontal que pasa por el valor ponderado global. Si no hay sesgo de publicación, aparece la forma de un embudo típico (simetría con relación a la línea dibujada).

Heterogeneidad (Heterogeneity). Significa que hay variabilidad estadística entre los estudios que se combinan. A veces se hace una distinción entre la heterogeneidad estadística (diferencias en los efectos reportados), heterogeneidad metodológica (diferencias en el diseño de los estudios) y heterogeneidad clínica (diferencias entre los estudios referidas a características clave de los participantes, a intervenciones o a medidas de resultado). Los tests estadísticos de heterogeneidad se utilizan para valorar si la variabilidad clínica en los resultados de los estudios (la magnitud de los efectos) es mayor que aquella que se esperaría hubiera ocurrido por azar.

Homogeneidad (Homogeneity). En las revisiones sistemáticas, se refiere al grado en el que los resultados de los estudios incluidos son similares. Puede hacer referencia a los participantes en el estudio, intervención y medidas de resultado. Se habla de ho-

mogeneidad cuando los resultados no varían más de lo que se esperaría por la influencia del azar.

Inverso de la varianza (Inverse of variance). Es el tipo más común de ponderación, utilizado para combinar diferentes estudios individuales en un estimador ponderado. Se puede aplicar a muchas medidas de magnitud del efecto.

Magnitud del efecto (Effect size). Es el estimador estandarizado no escalar de la relación entre una exposición y un efecto. En sentido general, este término se aplica a cualquier medida de la diferencia en el resultado entre los grupos de estudio; de tal manera que el riesgo relativo, la razón de momios y la diferencia de riesgos son magnitudes de efecto. Cuando la magnitud del efecto se aplica a mediciones de variables continuas (como la media), su estimador común más usado es la diferencia media estandarizada, que se calcula como la diferencia de medias, dividida por la desviación estándar. Esto es sobre todo útil cuando no hay una medida común a todos los estudios.

Metaanálisis (Meta-analysis, overview). Es la identificación sistemática, valoración, síntesis y, si es pertinente, la agregación estadística de todos los estudios sobre el mismo tema, siguiendo un método explícito y predeterminado.

Meta-regresión (Meta-regression). Colección de métodos estadísticos (regresión lineal ponderada o no, regresión logística) para valorar la heterogeneidad, en los que la magnitud del efecto se enfrenta a una o más covariables. Es un procedimiento que se puede utilizar cuando el número de estudios no es muy abundante (< 20).

Método de DerSimonian-Laird (DerSimonian-Laird's method). Fue el primer modelo de efectos aleatorios y se aplicó a un modelo aditivo (diferencia de riesgos), ponderado por el inverso de la varianza.²²

Método de Mantel-Haenszel (Mantel-Haenszel's method). Método estadístico de combinación de riesgos relativos y razón de momios de estudios individuales, que se escribió para combinar estratos de un mismo estudio. Hay que conocer la distribución cruda de los datos, por lo que suele ser apropiado para los estudios experimentales con asignación aleatoria, pero si los autores de los estudios originales proporcionan la información necesaria, puede aplicarse a cualquier tipo de estudio.

Método de Peto (Peto's method). Método estadístico para combinar estudios individuales, deri-

vado del método de Mantel-Haenszel, en el que los efectos observados en el grupo índice se comparan con los esperados, ponderando por la varianza. Debería usarse sólo cuando el tamaño de muestra de las ramas de un estudio es similar y cuando la magnitud de efecto es próxima al valor nulo; en otras situaciones habitualmente da un resultado sesgado.

Modelo de efectos aleatorios (Random effects model). Método de combinación de magnitudes de efectos individuales, en el que la heterogeneidad se incorpora al estimador resumen, mediante la inclusión de un componente de la variabilidad entre los estudios. Supone que la muestra de estudios incluidos en el análisis, se extrae de una población de estudios. Este modelo no supone homogeneidad en las magnitudes de efecto de los estudios que se combinan; esto es, cada muestra de estudio tiene una verdadera magnitud de efecto. No hay acuerdo sobre si este modelo es más apropiado que el de efectos fijos para combinar estudios. Se recomienda cuando el número de estudios es pequeño, siempre y cuando no haya sesgo de publicación, ya que es un método que da más importancia a los estudios pequeños.

Modelo de efectos fijos (Fixed effects model). Cualquier modelo estadístico en el que se asume homogeneidad del efecto en los diferentes estudios que se combinan; esto es, en el que la verdadera magnitud del efecto tiene un valor común y real para todos los estudios. En el estimador ponderado sólo se considera la varianza de cada estudio individual.

Número necesario a tratar, NNT (Number needed to treat). Es una medida de la eficacia de un tratamiento. Es el número de personas que se necesitaría tratar con un tratamiento específico para producir, o evitar, una ocurrencia adicional de un evento determinado.

Número de pacientes necesario para dañar (NNH). Se utiliza cuando una intervención aumenta el riesgo de un evento adverso. Similar al concepto de NNT, pero cuando el tratamiento experimental provoca efectos adversos. Es el número de pacientes que necesitamos tratar para obtener un evento: a cuántos pacientes sometidos al tratamiento experimental hay que tratar (comparado con el tratamiento control) para provocar un evento adverso. Se calcula como la inversa del aumento de riesgo absoluto: $1/$ (incidencia de evento adverso en el gru-

po experimental-incidencia del evento adverso en el grupo control).

Odds. Es un modo diferente de expresar la probabilidad. Cuanto más probable es que una cosa ocurra, respecto de que no ocurra (para odds pre prueba y odds post prueba véase probabilidad) .

Odds ratio/Razón de Odds (Odds ratio, OR). Es una medida de la eficacia de un tratamiento. Si es igual a 1, el efecto del tratamiento no es distinto del efecto del control. Si el OR es mayor (o menor) que 1, el efecto del tratamiento es mayor (o menor) que el del control. Nótese que el efecto que se está midiendo puede ser adverso (ej: muerte, discapacidad) o deseable (ej: dejar de fumar).

Peso (Weighting). En la influencia dada a cada estudio individual en el análisis combinado, hay diferentes tipos de ponderaciones: inverso de la varianza, por la distribución de los datos (Mantel-Haenszel), por la varianza de los casos esperados en la referencia (Peto), etcétera.

Placebo (Placebo). Sustancia inactiva o tratamiento dado para satisfacer las expectativas del paciente. En algunos ensayos clínicos controlados, el placebo está preparado de tal forma que lo hace indistinguible del tratamiento a estudio. Se da al grupo control para que permita realizar comparaciones y poner de relieve el efecto del tratamiento en investigación.

Poder estadístico o potencia estadística (Statistical power). Probabilidad de demostrar un efecto cuando realmente existe. Un estudio debe tener un diseño apropiado para que, en el caso de existir diferencias entre los grupos comparados, éstas puedan ser demostradas. Cuando nos encontramos ante un resultado negativo (no diferencia) debemos preguntarnos: ¿las poblaciones realmente no son diferentes, o el estudio no tuvo la capacidad suficiente como para demostrarlo? Esta capacidad del estudio para detectar diferencias cuando éstas existen se denomina poder estadístico, y está principalmente relacionado con el número de participantes en el estudio, el número de eventos y el tiempo de seguimiento.

Precisión (Precision). Grado en que un instrumento de medida o un estadístico produce los mismos resultados al aplicarse sobre la misma magnitud (instrumentos) o población (estadísticos). La precisión de un estadístico se estima por el intervalo de confianza.

Problema de la extracción de archivos (File drawer problem). Se emplea para referirse al número de estudios con resultados no significativos estadísticamente ($p > 0.05$) que no se han publicado.

QALY (Quality adjusted life year). Unidad de resultados en salud que se ajusta por ganancia o pérdida de años de vida tras una intervención sanitaria y según la calidad de vida durante esos mismos años. Puede utilizarse como unidad para comparar diferentes intervenciones sanitarias que analicen coste-utilidad.

Reducción relativa del riesgo, RRR (Relative risk reduction). Es el cociente entre la RAR y el riesgo en el grupo control. Generalmente se expresa en porcentaje. También llamado fracción atribuible.

Riesgo (Risk). Probabilidad de ocurrencia de un evento concreto. Se estima por la proporción de individuos en los que se observa el evento. Si en un grupo de 100 se observan 12 eventos, el riesgo estimado es 0.12 ó 12%.

Riesgo atribuible (Attributable risk). Es la diferencia entre el riesgo en el grupo control y riesgo en el grupo tratado.

Riesgo relativo, (RR) (Risk ratio o relative risk). Es el cociente entre el riesgo en el grupo tratado y el riesgo en el grupo control. Es una medida de la eficacia de un tratamiento. Si es igual a 1, el efecto del tratamiento no es distinto del efecto del control. Si el RR es mayor (o menor) que 1, el efecto del tratamiento es mayor (o menor) que el del control.

Revisión sistemática (Systematic review). Síntesis de los resultados de varios estudios primarios mediante técnica que limitan los sesgos y el error aleatorio. Estas técnicas incluyen la búsqueda de todos los estudios potencialmente relevantes y el uso de criterios explícitos y fiables en la selección de las investigaciones. La revisión sistemática cualitativa, resume las investigaciones primarias u originales sin combinación estadística (metaanálisis cualitativo). La revisión sistemática cuantitativa es sinónimo de metaanálisis.

Sesgo (Bias). Cualquier error sistemático en el diseño, desarrollo o análisis de un estudio que pueda producir desviaciones en la estimación del efecto, teniendo como consecuencia una infravaloración o sobrevaloración del verdadero efecto de una intervención.

Sesgo de búsqueda (Bias search). Error sistemático introducido cuando la búsqueda se centra en una sola base de datos (normalmente Medline). Las revistas escritas en inglés están más representadas en Medline; además, las revistas de un país (y de países vecinos o de similar idioma o cultura) donde se confecciona la base de datos, también están proporcionalmente más representadas. Es recomendable consultar más de una base de datos, junto con una búsqueda manual de las referencias localizadas en cada estudio.

Sesgo de idioma (Bias language). Error sistemático introducido cuando la búsqueda de estudios potenciales de un metaanálisis se centra en un solo idioma, en general el inglés. Es posible que los resultados significativos alcanzados en países de habla no inglesa, se publiquen más frecuentemente en inglés que en su idioma nativo, por la mayor difusión alcanzada.

Sesgo de publicación (Bias publication). Sesgo introducido cuando los estudios publicados no representan adecuadamente todos los realizados sobre un tema específico. Puede deberse a muchos hechos, aunque el mejor conocido es la tendencia a publicar resultados estadísticamente significativos ($p < 0.05$) o clínicamente relevantes (magnitud del efecto alta, aunque no significativa). Otras variables que influyen en el sesgo de publicación son el tamaño de muestra (mayor sesgo en estudios pequeños), tipo de diseño (menor en los ensayos clínicos con asignación aleatoria), financiación, conflicto de intereses, prejuicio frente a una asociación y el patrocinio.

Tamaño muestral (Sample size). Número total de sujetos que se necesitan para realizar un estudio, incluyendo todos los grupos de intervención. Se calcula usando una fórmula estadística basada en el error tipo I y tipo II, la diferencia clínicamente relevante entre los dos grupos y la varianza asociada. Si se detecta una pequeña diferencia, se necesitará un gran tamaño muestral.

Validez (Validity). Solidez o rigor de un estudio, en relación con el grado de aproximación a la verdad de sus resultados. Un estudio es válido si el modo en que ha sido diseñado y realizado hace que los resultados no estén sesgados; es decir, nos da una verdadera estimación de la efectividad clínica.

Validez externa (External validity). Grado en que los resultados de un estudio son generalizables a otros individuos.

Validez interna (Internal validity). Grado en el que los efectos observados son verdaderos para las personas del estudio.

Valor P (P value). La probabilidad (que va de cero a uno) de que los resultados observados en un estudio o los resultados más extremos que los observados, puedan haber ocurrido por azar. En un metaanálisis, el valor P para el efecto global evalúa la significación estadística global de la diferencia entre los grupos tratamiento y control, mientras que el valor P para los estudios de heterogeneidad objetiva, la significación estadística estriba en diferencias entre los efectos observados en cada estudio.

REFERENCIAS

1. Heinrichs KI. Problem-Based Learning in Entry-Level Athletic Training Professional-Education Programs: A Model for Developing Critical-Thinking and Decision-Making Skills. *J Athl Train* 2002; 37(4 Suppl): S189-S198.
2. Sackett D. Medicina basada en la evidencia: Cómo ejercer y enseñar la MBE. Churchill Livingstone, 1997.
3. Chalkidou K, Walley T, Culyer A, Littlejohns P, Hoy A. Evidence-informed evidence-making. *J Health Serv Res Policy* 2008; 13: 167-73.
4. Franco G, Monduzzi G. Experimental validation of the Evidence-Based Occupational Health paradigm and of the PICO model in the decision making process applied by occupational health physicians. *Med Lav* 2004; 95: 423-30.
5. Pólya G. How to Solve it: A New Aspect of Mathematical Method. 2nd ed., Princeton University Press, 1957, ISBN 0-691-08097-6.
6. Perera R, Heneghan C. Interpreting meta-analysis in systematic reviews. *Evid Based Med* 2008; 13: 67-9.
7. Shrier I, Boivin JF, Platt RW, Steele RJ, Brophy JM, Carnevale F et al. The interpretation of systematic reviews with meta-analyses: an objective or subjective process? *BMC Med Inform Decis Mak* 2008; 8: 19.
8. Tseng TY, Dahm P, Poolman RW, Preminger GM, Canales BJ, Montori VM. How to Use a Systematic Literature Review and Meta-Analysis. *J Urol* 2008 Aug 14. [Epub ahead of print]
9. Zwahlen M, Renehan A, Egger M. Meta-analysis in medical research: potentials and limitations. *Urol Oncol* 2008; 26: 320-9.
10. Ruano-Raviña A, Figueiras A, Barros-Dios JM. El metaanálisis a debate. *Med Clin (Barc)* 2002; 119: 435-439.
11. Jefferson T, Rudin M, Brodny Folse S, Davidoff F. Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database Syst Rev.* 2007 Apr 18; (2): MR000016.
12. Guyatt GH, Mills EJ, Elbourne D. In the era of systematic reviews, does the size of an individual trial still matter. *PLoS Med* 2008 Jan 3; 5(1): e4.

13. Delaney A, Bagshaw SM, Ferland A, Laupland K, Manns B, Doig C. The quality of reports of critical care meta-analyses in the Cochrane Database of Systematic Reviews: an independent appraisal. *Crit Care Med* 2007; 35: 589-94.
14. Lundh A, Gøtzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. *BMC Med Res Methodol* 2008; 21; 8: 22.
15. Tricco AC, Tetzlaff J, Sampson M, Fergusson D, Cogo E, Horsley T, Moher D. Few systematic reviews exist documenting the extent of bias: a systematic review. *J Clin Epidemiol* 2008; 61: 422-34.
16. Shrier I, Boivin JF, Steele RJ, Platt RW, Furlan A, Kakuma R, Brophy J, Rossignol M. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol.* 2007; 166: 1203-9.
17. Shrier I, Boivin JF, Platt RW, Steele RJ, Brophy JM, Carnevale F et al. The interpretation of systematic reviews with meta-analysis: an objective or subjective process? *BMC Med Inform Decis Mak* 2008; 8: 19.
18. Moher D, Tsertsvadze A, Tricco AC, Eccles M, Grimshaw J, Sampson M, Barrowman N. When and how to update systematic reviews. *Cochrane Database Syst Rev* 2008 Jan 23; (1): MR000023.

Correspondencia:

Dr. Héctor A. Baptista González.
Montes Urales 800, Lomas Virreyes,
Delegación Miguel Hidalgo.
11000 México, D.F.
baptista@infosel.net.mx.