

Revisiones Sistemáticas: Su Metodología y Aplicaciones

■
VOLANDA TORRES DE GALVIS¹ RUBEN DARÍO MANRIQUE HERNÁNDEZ²

RESUMEN

Se revisan las características y ventajas de las revisiones sistemáticas, haciendo énfasis en su utilidad y en los pasos para su preparación y actualización. La formulación del problema, los aspectos claves en la formulación de la pregunta, las estrategias de búsqueda para la localización y selección de los estudios primarios, la evaluación crítica de los estudios, cuidando su validez a través del control de los posibles sesgos; análisis y presentación de resultados y finalmente se presentan las guías para la lectura crítica de las RS y la utilización de sus resultados.

PALABRAS CLAVES

Revisión sistemática
Meta-análisis
Estrategias de búsqueda
Sesgos de publicación
Lectura crítica

¹ Jefe División de Investigaciones Facultad de Medicina CES. E-mail ytorres@ces.edu.co

² Profesor División de Investigaciones Facultad de Medicina CES. E-mail rmanrique@ces.edu.co

SUMARY

The characteristics and advantages of systematic review are reviewed, we are doing emphasis in their utility and the steps for their preparation and update. The formulation of the problem, key aspects in the formulation of the question, the search strategy for location and selection of primary studies, the critical evaluation of studies, taking care of its validity through the control of possible bias; analyses and presentation of results and finally appear guides for the critical appraisal of RS and the use of their results.

KEY WORKS

Systematic review
Meta-analysis
Publication bias
Search Strategy
Critical appraisal

INTRODUCCIÓN

- La toma de una decisión siempre está acompañada de cierto grado de incertidumbre. Cada vez se exige más que las decisiones clínicas estén basadas en evidencias sólidas procedentes de estudios clínico-epidemiológicos. (Martí J, Bonfil X, Mejía R y Tohá D. 1999)
- Con este artículo la revista CES Medicina, inicia la publicación de una serie sobre metodología de la investigación y Medicina Basada en la Evidencia como un aporte a quienes generan o utilizan la investigación científica para el avance en el desarrollo de las ciencias de la

salud y el mejoramiento de la atención de nuestra población.

Desde siempre los médicos han buscado basar las decisiones y acciones en la mejor evidencia posible, ¿porqué la necesidad de adoptar nuevas terminologías y nuevas técnicas de resumir la generación de conocimientos y evaluar su validez?. La cantidad de información de que se dispone es abrumadora, y además crece a un ritmo vertiginoso. Es imposible mantenerse al día en todas las áreas de una especialidad médica, e incluso a veces en las de una única sub-especialidad.

Los métodos tradicionales de resumen de la evidencia disponible, libros y revisiones tradicionales no sistemáticas, si bien poseen su valor, tienen problemas que limitan su utilidad:

- Suelen ser excesivamente generales para problemas concretos.
- Con frecuencia no incorporan hallazgos recientes.
- No resumen la evidencia de forma sistemática, es decir, el lector desconoce qué criterios ha seguido el autor para dar más énfasis a unos resultados o a otros.

Por tanto los clínicos y los administradores de salud necesitan herramientas para sintetizar de forma sistemática la evidencia disponible acerca de un problema clínico concreto. Esta síntesis, debe realizarse con criterios explícitos y reproducibles, las revisiones sistemáticas, los meta-análisis y las guías de atención clínica basadas en la evidencia ayudan a resolver estos problemas.

Dentro de la investigación por revisión de publicaciones se reconocen dos tipos diferentes: sistemática y no sistemática o «revisión narrativa».

La diferencia esencial entre revisión sistemática y no sistemática está dada, porque en esta última falta un protocolo que defina los pasos que siguió el revisor para actualizar el tópico. Por ejemplo ¿qué estrategia se utilizó para la búsqueda bibliográfica?, cómo se hizo la selección de estudios a ser incluidos en la revisión y/o qué métodos se usaron para determinar la validez de los mismos?

A su vez dentro de la revisión sistemática existen dos formas: «cuantitativa o meta-análisis» y «cualitativa u overview». Las diferencias están dadas fundamentalmente por el uso de métodos estadísticos, en el meta-análisis, que permiten la combinación y análisis cuantitativos de los resultados obtenidos en cada estudio.

Las revisiones sistemáticas (RS) se pueden considerar como «estudios de estudios», son procedimientos científicos útiles para resumir, valorar y comunicar los resultados de lo que de otra manera serían acumulaciones inmanejables de información científica. Para ello, resumen y sintetizan los resultados de múltiples «investigaciones primarias» utilizando estrategias que limitan los sesgos y los errores aleatorios

Las RS ayudan a estar al día sobre el mejor conocimiento médico sintetizando la información científica. Su utilización es creciente, tanto para fundamentar mejor las decisiones médicas como para perfilar programas de investigación o de política sanitaria. Las RS refuerzan los vínculos entre la mejor investigación médica y la atención óptima a la salud de los pacientes. (Chalmers I, Altman DG, 1995)

Las RS son necesarias para integrar eficientemente toda la información válida y proporcionar una base para tomar decisiones de

manera racional. Establecen cuándo los efectos de la atención de salud son consistentes y cuándo pueden aplicarse los resultados de la investigación a los diferentes grupos de población, ámbitos y diferencias en el tratamiento (por ejemplo, en cuanto a dosis), y determinan dónde pueden variar significativamente los efectos.

Principales características de las RS

- La pregunta de investigación que se pretende revisar se plantea de forma explícita, como si se tratara de un estudio original.
- Se hace un esfuerzo sistemático por localizar toda la evidencia disponible, no sólo los estudios que aparecen en las bases de datos bibliográficas más extendidas.
- Todos los estudios recuperados se revisan según criterios comunes establecidos de antemano, y se plantean criterios objetivos para decidir qué estudios tienen información apropiada para responder a la pregunta de investigación.
- Los resultados de todos los estudios se presentan de forma objetiva e imparcial y se combinan con métodos estadísticos que dan más peso a los estudios más precisos, obteniendo medidas de resumen cuantitativas.

El uso de métodos explícitos y sistemáticos en las revisiones limita el sesgo (errores sistemáticos) y reduce los errores aleatorios (errores simples), proporcionando así resultados más confiables sobre los cuales sacar conclusiones y tomar decisiones. El meta-análisis, uso de los métodos estadísticos para resumir los resultados de estudios independientes entre sí, puede proporcionar estimaciones más precisas que aquellas derivadas de los estudios individuales incluidos en una revisión. (Dickersin K, Min YI. 1993)

El más amplio reconocimiento de la función clave que tienen las revisiones en la síntesis y difusión de los resultados de la investigación ha ayudado a valorar su validez. En los años setenta y principio de los ochenta, psicólogos y profesionales de las ciencias sociales hicieron hincapié en los pasos sistemáticos necesarios para reducir al mínimo los sesgos y los errores aleatorios en las revisiones de estudios científicos. Pero sólo hasta fines de los años ochenta, los diferentes sectores señalaron la deficiente calidad científica de muchos artículos de revisión sobre temas de salud.

A partir de entonces el reconocimiento de la necesidad de las RS en este campo ha crecido rápidamente y sigue extendiéndose, tal y como queda reflejado por el número creciente de artículos acerca de métodos de revisión y del número de revisiones sistemáticas publicadas en revistas de salud.

Ventajas de las RS

La consideración sistemática de toda la información disponible y la necesidad de evaluar y extraer la información de cada estudio de forma objetiva son, tal vez, las principales ventajas de las RS. De esta forma, se minimiza la posibilidad de olvidar estudios importantes no identificados, así como la posibilidad de sesgos que pueden aparecer en una selección subjetiva de la bibliografía.

Además, la obtención de resultados globales cuantitativos permite conocer de manera objetiva la magnitud del efecto de las intervenciones y puede facilitar la toma de decisiones clínicas. No es de extrañar, por tanto, que las RS y los meta-análisis sean piedras angulares de la Medicina

Basada en la Evidencia, y que los grupos que trabajan en este campo como la Colaboración Cochrane, hayan adoptado ampliamente estas técnicas.

Las RS permiten aumentar la precisión de las estimaciones, con mayor capacidad para detectar efectos que los estudios individuales, y evaluar las discrepancias entre los distintos estudios primarios, sugiriendo hipótesis explicativas de estas diferencias.

Las RS han puesto de relieve la importancia del sesgo de publicación en la interpretación de la literatura médica. El término «sesgo de publicación» se refiere al grado en el que los artículos científicos publicados no reflejan fielmente las investigaciones realizadas, ya que aquellos con resultados estadísticamente significativos tienen mayor probabilidad de ser enviados a publicación, de ser aceptados, y de aparecer en revistas con mayor difusión e impacto. En otras palabras, los estudios publicados son una muestra sesgada de todos los estudios, con una tendencia a exagerar la eficacia de las intervenciones. (Dickersin K, Min YI. 1993; Dickersin K, Scherer R, Lefebvre C. 1994)

Frecuentemente, los ensayos que no demuestran que una intervención es eficaz tienen menos probabilidades de ser publicados que aquellos otros que muestran un efecto favorable. Esto significa que las revisiones sistemáticas que dejan de incluir ensayos inéditos quizá estén sesgadas por una sobrestimación de la eficacia de una intervención. (Schulz KF, Chalmers I, Hayes RJ, Altman DG.)

Aun cuando en este artículo se hace énfasis en la metodología utilizada por la Colaboración Cochrane la cual se centra las revisiones sistemáticas ensayos clínicos controlados con asig-

nación aleatoria (RCTs: *randomized controlled trials*) por tener mayor posibilidad de proporcionar información más confiable que otras fuentes de evidencia *sobre los efectos diferenciales de las distintas alternativas de tratamiento o prevención*, no puede olvidarse que las RS de otros tipos de estudios, también pueden ayudar a quienes desean tomar las mejores decisiones, en particular sobre aquellos campos en los cuales no se pueden hacer RCTs, por cuestiones de índole ética, como es el caso de evidencias sobre factores de riesgo.

Los principios básicos de revisión de las investigaciones son los mismos, cualquiera que sea el tipo de evidencia que se someta a examen.

Aunque las RS se centran principalmente en RCTs, debe quedar claro que éstas se pueden ocupar de temas específicos o preguntas que lleven a la revisión de otras fuentes de evidencia con metodologías diferentes cuando resulten pertinentes.

Pasos en la Preparación o Actualización de una RS

Como en cualquier proyecto de investigación, los métodos que se usen deben establecerse de antemano. Puesto que las revisiones son análisis retrospectivos de datos que han sido recogidos por otros, es importante que el proceso esté bien definido y tan riguroso como sea posible.

Los siguientes son los pasos que se requieren (según la metodología seguida por los grupos de la Colaboración Cochrane) para la preparación o actualización de una revisión sistemática:

- Formulación del problema.
- Localización y selección de los estudios.

- Evaluación crítica de los estudios.
- Obtención de los datos.
- Análisis y presentación de los resultados.
- Interpretación de los resultados.
- Mejora y actualización de las revisiones.

ANTECEDENTES DE LA REVISIÓN

El protocolo debe comenzar con una breve síntesis de los aspectos que han motivado y justificado la revisión. En general, las revisiones sistemáticas pueden tener como justificación varios factores. Por ejemplo, una revisión puede llevarse a cabo en un intento de resolver evidencias contradictorias, o bien puede tratar de contestar preguntas cuya respuesta sea incierta, o bien explicar variaciones de la práctica clínica.

Pueden estar orientadas a proporcionar resúmenes fiables, objetivos y actualizados de aquello que se conoce y aquello que se ignora acerca de los efectos de las diferentes formas de atención médica y ayudar a la toma de decisiones que sean prácticas

Consideraciones a tener en cuenta:

- La formulación de preguntas debe abordar las diferentes alternativas (opciones prácticas) a las que nos enfrentamos cuando tomamos decisiones.
- Deben tener en cuenta los resultados (**outcomes**) que sean significativos o relevantes para tomar las decisiones.
- Los métodos a utilizar en una revisión deben ser los que permitan optimizar las probabilidades de que los resultados proporcionen la

mejor evidencia disponible sobre la cual basar nuestras decisiones.

- Es importante que los lectores puedan conocer cuándo no existen evidencias confiables, o si éstas no existen en relación a variables de resultado específicas que sean de importancia para la toma las decisiones.
- Aún cuando no exista otra mejor evidencia disponible no resulta de utilidad incluir en una revisión aquella evidencia que tiene alta probabilidad de estar sesgada.
- De igual manera, no tiene sentido centrarse en variables de resultado superficiales o triviales simplemente por el hecho de que éstas sean las que los investigadores han seleccionado para evaluar los efectos de la intervención.
- Los estudios considerados por la revisión no deben quedar restringidos por un criterio de nacionalidad o de idioma sin una razón de peso que lo justifique.
- Los resultados deben ser presentados de tal forma que su aplicabilidad en diferentes contextos pueda ser evaluada por los responsables de la toma de decisiones.
- *El objetivo primario de una RS debe ser resumir y ayudar a las personas a comprender la evidencia. Los revisores deben ser cuidadosos en no imponer a otros sus propios valores y preferencias al tratar de responder las preguntas que plantean.*

Las preguntas bien formuladas generalmente no aparecen espontáneamente, surgen en el contexto de un cuerpo de conocimiento ya formado. Este contexto debe ser abordado en la sección de antecedentes de la revisión. Debe mencionarse cualquier revisión sistemática sobre la misma cuestión que haya sido publicada previamente o cualquier controversia existente al respecto.

FORMULACIÓN DEL PROBLEMA

- Las preguntas mal enfocadas conducen a decisiones poco claras y confusas acerca de qué estudios conviene incluir en la revisión y cómo resumir sus datos.

Del mismo modo que en cualquier proyecto de investigación, la primera y más importante decisión, es determinar su enfoque. Esto se consigue mejor planteando preguntas claramente delimitadas. Resultan esenciales a la hora de determinar la estructura de una revisión. Concretamente, guiarán gran parte del proceso de revisión como las estrategias para la localización y la selección de los estudios o de los datos, para valorar críticamente su relevancia y validez y para analizar la variación observada entre sus resultados.

Además de guiar el proceso de revisión, las preguntas y los objetivos de una revisión sirven también a los lectores en su valoración inicial de la relevancia de la misma.

Componentes Claves en una Pregunta

Una pregunta claramente definida debe especificar:

- Tipo de personas (participantes)
- Tipo de intervenciones o de exposiciones
- Resultados que son de interés para la revisión.
- Además, debe especificarse el tipo de estudios que son pertinentes para tratar de responder a la pregunta planteada.

¿Qué tipo de personas?

A menudo, resulta de utilidad considerar en dos fases el tipo de personas objeto de interés para la revisión, identificando la población y el ámbito de interés. En primer lugar, definiendo la *enfermedad o condición* que se considere de interés, haciendo suficientemente explícitos los criterios diagnósticos para determinar la presencia de esa enfermedad o en qué condiciones podría desarrollarse. Esto implica tener que decidir si uno está interesado en un grupo específico de la población, en función de factores tales como la edad, sexo, grado de instrucción o presencia de una enfermedad particular

¿Tipo de comparaciones?

El siguiente componente clave de una pregunta bien planteada es especificar las intervenciones que son de interés. También es importante definir el tipo de grupos control que resulten aceptables para la revisión. Debe considerarse si las personas en el grupo control quizá pudieran estar sometidas a otras intervenciones distintas al placebo y si aquellas intervenciones se superponen de alguna manera con la intervención activa que está siendo evaluada.

¿Qué resultados?

El tercer componente clave es la especificación de aquellos resultados concretos que sean de interés. ¿Es necesario definir unos criterios explícitos para establecer la presencia de aquellos resultados?

En general, las RS deben incluir todos aquellos resultados que puedan ser significativos para las personas que necesitan tomar alguna decisión acerca del problema del que se ocupa la revisión.

A veces también resulta importante especificar los resultados que sean más trascendentes para la toma de decisiones, aún cuando sea improbable que se encuentren esos datos.

Todos los resultados que se consideren importantes deben estar incluidos en las revisiones, mientras que los más triviales no deberán incluirse. Antes de excluir de una revisión un estudio que cumple los requisitos de selección pero que no presenta los resultados de modo adecuado, para la revisión, vale la pena considerar la posibilidad de tratar de obtener la información que se necesita directamente de los investigadores para complementar lo que haya sido publicado.

¿Qué tipo de estudios?

Algunos tipos de estudios son superiores a otros para tratar responder a determinadas preguntas. Se recomienda que los ensayos clínicos aleatorizados sean condición *sine qua non* para tratar asuntos referentes a la eficacia terapéutica, mientras que otros tipos de diseño resultan más apropiados para abordar otro tipo de preguntas.

Por ejemplo las preguntas relacionadas con la etiología o los factores de riesgo deben ser abordadas por estudios de cohortes o de casos y controles.

Las preguntas referentes a las estrategias diagnósticas o de revisión selectiva (*screening*) pueden ser abordadas de forma óptima mediante ensayos clínicos aleatorizados, pero las preguntas sobre la exactitud de las pruebas diagnósticas se responden mejor mediante estudios transversales (*cross-sectional*).

- Se debe considerar en cada caso qué tipo de diseño de los estudios puede proporcionarnos

datos fiables con los cuales responder a las preguntas planteadas.

- La revisión de estudios que proporcionan datos poco fiables con los que responder a una pregunta, aparte de ser una pérdida de tiempo, puede conducir a conclusiones erróneas.

Utilización de los componentes clave de la pregunta para identificar y seleccionar los estudios

- Una vez que se tiene la pregunta de investigación bien planteada, se debe determinar qué componentes claves se tendrán en cuenta para definir la estrategia de búsqueda inicial

En general, los términos o palabras clave utilizados en una búsqueda consisten en la condición o enfermedad que se busca, la intervención, exposición, o la prueba diagnóstica a evaluar. Aunque quizá se esté específicamente interesado en un ámbito determinado, los estudios no suelen estar indexados según este parámetro en muchas de las bases de datos electrónicas.

- Cualquiera que sea la estrategia de búsqueda utilizada, será necesario revisar varios artículos y decidir cuáles son relevantes y cuáles no.

La formulación de la pregunta en función del tipo de participantes, intervenciones, resultados y diseño de los estudios considerados obligará a especificar los criterios que se usarán para seleccionar los estudios. Sin embargo, se requiere de una mayor clarificación de los criterios de selección para desarrollar reglas de decisión que sean sensibles y reproducibles.

Los revisores registrarán aquellos datos relevantes en relación con los componentes clave de la pregunta en cada uno de los estudios individuales. Por lo tanto, las preguntas bien formuladas están relacionadas directamente al proceso de recopilación de datos porque guían:

- A la hora de establecer los criterios que se utilizarán para seleccionar los estudios adecuados para la revisión
- Determinan qué datos deben ser extraídos de los estudios que satisfacen aquellos criterios de selección.

La validez de las revisiones definidas en términos muy amplios puede ser criticada, en particular cuando existe sólida evidencia biológica o sociológica que sugiere que las diversas formas de intervenciones se comportan de modos muy diversos, o que las varias definiciones de la condición de interés o del entorno se asocian con una evolución y unos resultados marcadamente diferentes. Está bien mezclar manzanas y naranjas si la pregunta versa sobre fruta en general, pero no si la pregunta es, por ejemplo, acerca de la vitamina C cuando uno conoce de antemano que manzanas y naranjas son diferentes en lo que se refiere a esta vitamina.

LOCALIZACIÓN Y SELECCIÓN DE LOS ESTUDIOS

La búsqueda de los estudios global y no sesgada es una de las diferencias clave entre *una revisión sistemática y una revisión tradicional*. Si bien las bases de datos electrónicas -como MEDLINE- son potentes herramientas para la localización de

estudios, sólo pueden identificarse a partir de MEDLINE entre un 30 y un 80% de todos los ensayos controlados y aleatorizados que han sido publicados, dependiendo también del área o tema específico. Además, en MEDLINE las referencias de estudios escritos en lengua no inglesa son poco representativas y sólo están indexados los artículos publicados, lo que constituye un potencial sesgo de publicación e idiomático (Begg CB, Berlin JA. 1998). Esto significa que para evitar estos sesgos y asegurar que todos los estudios relevantes queden incluidos en una revisión es necesario un enfoque sistemático y el uso de múltiples fuentes de información.

De acuerdo con los principios que guían el desarrollo de una RS, cada grupo de revisión debe procurar satisfacer sus necesidades específicas de identificación de los estudios, de una manera lo más amplia posible, de protección contra los sesgos y de reducción al mínimo de la duplicación innecesaria de esfuerzos.

La fuente principal de estudios para las revisiones son los registros de ensayos de los grupos de revisión Cochrane. Así se asegura, en parte, que todos los estudios relevantes en el registro queden incluidos en una revisión, pero se debe considerar la posibilidad de usar estrategias adicionales para asegurar aún más que todos los estudios pertinentes queden registrados.

Al desarrollar una estrategia de búsqueda es importante evitar la duplicación innecesaria de esfuerzos, sacar el máximo provecho a las bases de datos, y contribuir a familiarizarse y a usar de manera más eficaz otras estrategias de búsqueda y bases de datos alternativas.

Se deben revisar las listas de referencias bibliográficas de los artículos válidos identificados. En

general, éste es un medio eficaz de identificar otros estudios, aunque pueda ser lento. Los artículos adicionales que se identifiquen como potencialmente relevantes deben obtenerse y ser evaluados para su posible inclusión. Debe tenerse en cuenta el potencial sesgo de referencia, tendencia subjetiva a citar preferentemente estudios que apoyan los propios criterios.

Algunas de las fuentes de referencias de ensayos más obvias y convenientes son las revisiones anteriores sobre el mismo tema. Se deben obtener y examinar las copias de las revisiones publicadas con anterioridad sobre el tema de interés para tener las referencias de los estudios originales. Este proceso de seguir las referencias de un artículo a otro se denomina *pearling* en inglés, método de la ascendencia o de concatenación de citas.

MEDLINE, EMBASE y otras bases de datos bibliográficas también pueden usarse para identificar artículos de revisión. En MEDLINE, los artículos de revisión están indexados en el apartado *Publication Type* como *Reviews*. Si se encuentra un gran número de revisiones, a menudo es útil limitar la búsqueda a las codificadas como «*Meta-analysis*» o «*Review, Academic*» en *Publication Type*.

Comunicación personal

Los colegas pueden ser una fuente importante de información de ensayos recientes que no hayan sido publicados todavía o de ensayos más antiguos que nunca llegaron a publicarse. Estos canales informales de comunicación a veces pueden ser los únicos medios para identificar resultados no publicados y «literatura gris», tales como informes técnicos, actas de conferencias o tesis doctorales.

Registro Cochrane de Ensayos Controlados

El Registro Cochrane de Ensayos Controlados (CCTR) forma parte de The Cochrane Library y contiene referencias de los ensayos controlados aleatorios y ensayos clínicos controlados en salud. Estas referencias son aportadas por grupos de revisión y otras entidades que han identificado los ensayos a través de diversos mecanismos de búsqueda. Tanto el contenido como la indización del CCTR están evolucionando para mejorar la utilidad que puedan prestar a los revisores como fuente principal para la identificación de ensayos.

The Cochrane Library incluye la Base de Datos de York de Resúmenes de Revisiones de Eficacia (*Database of Abstracts of Reviews of Effectiveness-DARE*). La DARE proporciona información acerca de las revisiones publicadas anteriormente sobre los efectos de la atención sanitaria. Incluye.

Otras bases de datos bibliográficas

Otras bases de datos bibliográficas pueden ser, ocasionalmente, una fuente útil de información adicional para localizar ensayos para una revisión específica. Pueden usarse las bases de datos «*on line*» en Ciencias Médicas y Biológicas -una guía de 795 bases de datos- para localizar y evaluar la relevancia de otras bases de datos adicionales, así como la base de datos LILACS y otras fuentes de literatura científica en español.

Búsqueda manual

Para facilitar la identificación de todos los ensayos publicados, la Colaboración Cochrane ha propuesto, la estrategia de búsqueda manual. Por búsqueda manual se entiende el examen de cada número de una revista y la lectura, de manera

progresiva, de cada uno de sus títulos, resúmenes y artículos, hasta que sea posible determinar si el artículo es un ensayo controlado aleatorio (RCT), un ensayo clínico controlado (CCT) o un meta-análisis.

DESARROLLO DE UNA ESTRATEGIA DE BÚSQUEDA

En el protocolo debe describirse el proceso por el cual los estudios serán seleccionados para su inclusión en la revisión, es decir, la estrategia de búsqueda. Estos procedimientos, cuanto más rigurosos, permiten mejor la reducción de los errores y sesgos, pero también aumentan los recursos y el tiempo necesarios para preparar la revisión. La selección de los estudios para incluirlos en una revisión es un proceso de "embudo" que incluye múltiples etapas.

Cuando se revisan los resultados de una búsqueda electrónica, la primera etapa incluye la lectura detallada de los títulos y resúmenes para determinar si cada artículo reúne los requisitos predeterminados de elegibilidad.

En cada etapa del proceso de selección es preferible errar por exceso que por defecto, porque una vez un estudio ha quedado excluido del proceso de selección es improbable que se vuelva a considerar su inclusión. Los artículos dudosos que sean incluidos en una fase previa pueden excluirse en una etapa posterior, cuando se disponga de toda la información sobre el estudio.

Selección de estudios

Todos los artículos identificados deben ser posteriormente evaluados para ver si cumplen los

criterios de inclusión para la revisión. Los revisores deben decidir:

- Si uno o varios revisores evaluarán la relevancia de cada artículo.
- Si las decisiones en cuanto a la relevancia serán tomadas por especialistas en el tema, por no especialistas, o por ambos.
- Si los revisores que evalúan la relevancia de los estudios dispondrán de los nombres de los autores, instituciones, revistas donde se publicaron los estudios y resultados una vez se apliquen los criterios de inclusión.
- Cómo se resolverán las discrepancias si existe más de un revisor.

Las decisiones acerca de qué estudios incluir a menudo incorporan opiniones. Para ayudar a asegurar la reproducibilidad de estos juicios, es aconsejable que más de un revisor aplique los criterios de inclusión a todos los artículos evaluados y potencialmente relevantes. El número de personas que evalúan la relevancia de cada ensayo debe informarse en la sección de Métodos de la revisión.

Prueba Piloto

Para la mayoría de las revisiones puede ser útil la realización de una prueba piloto sobre la aplicación de los criterios de inclusión a una muestra de artículos (de diez a doce publicaciones, incluidas aquellas que se supone que son definitivamente elegibles, definitivamente no elegibles y dudosas). La prueba piloto puede usarse para pulir y clarificar los criterios de inclusión, entrenar a las personas que los aplicarán y asegurarse de que los criterios puedan ser aplicados sistemáticamente por más de una persona.

Cómo documentar una estrategia de búsqueda

La estrategia de búsqueda debe describirse con suficiente detalle como para que el proceso pueda reproducirse. La documentación de la estrategia de búsqueda debe incluir las principales fuentes de información utilizadas para identificar los estudios, la estrategia de búsqueda para cada base de datos consultada, las limitaciones temporales (por ejemplo, entre 1980 a 2001); y las limitaciones de idioma, si hubiese alguna (aunque deben ser evitadas en la medida de lo posible).

EVALUACIÓN CRÍTICA DE LOS ESTUDIOS

La evaluación crítica de los estudios individuales que se incluyen en las RS es necesaria para limitar los sesgos, formarse una idea de las potenciales comparaciones y guiar la interpretación de resultados. Los parámetros que requieren evaluación son aquellos relacionados con la aplicabilidad de los resultados, la validez de los estudios individuales y ciertas características del diseño que afectan la interpretación de los resultados.

Validez

En el contexto de una RS, la validez de un estudio se refiere a la capacidad de evitar los errores sistemáticos o sesgos mediante el diseño y la realización del estudio. Un concepto importante que no debe ser confundido con la validez es la precisión. La precisión es una medida del error aleatorio. Se refleja en el intervalo de confianza de la estimación del efecto de cada estudio y en el peso

dado a los resultados de cada estudio cuando se obtiene una estimación global del efecto (habitualmente una media ponderada) usando meta-análisis. A los resultados más precisos se les da más peso.

Aunque no haya diferencias en la validez, ni en los resultados de los estudios incluidos en una revisión, es importante evaluar críticamente todos los estudios. Por ejemplo: podría ocurrir que los resultados fueran consistentes en todos los estudios evaluados pero que los estudios fueran imperfectos.

Tipos de sesgo en los ensayos clínicos

En los estudios sobre los efectos una intervención se pueden presentar cuatro tipos de errores sistemáticos:

- De selección
- De información
- De realización (*performance bias*).
- De desgaste o pérdida (*attrition bias*) y
- De detección

Cuadro Nº 1. Tipos de Sesgos

Tipos de sesgo	Población diana	
	Asignación	
• SESGO DE SELECCIÓN (diferencias sistemáticas en los grupos a comparar)	Grupo de intervención ↓	Grupo de control ↓
• SESGO DE REALIZACIÓN (diferencias sistemáticas en la atención sanitaria proporcionada, con independencia de la intervención objeto de evaluación)	Expuestos a la intervención ↓	No expuestos a la intervención ↓
• SESGO DE DESGASTE (diferencias sistemáticas en las pérdidas de pacientes)	Seguimiento ↓	Seguimiento ↓
• SESGO DE DETECCIÓN (diferencias sistemáticas en la evaluación de los resultados puede ser debida a la no aplicación del ciego o doble ciego)	Medidas de resultado	Medidas de resultado

Sesgo de selección

- Uno de los sesgos más importantes que puede distorsionar las comparaciones entre tratamientos está relacionado con el proceso de obtención de los grupos a comparar.

En el diseño de ensayos clínicos es muy importante utilizar un método apropiado para prevenir la asignación sesgada a los grupos de tratamientos.

Sesgo de realización

- Con "sesgo de realización" (*performance bias*) se hace referencia a las diferencias sistemáticas en la atención proporcionada a los diferentes grupos de comparación de la intervención que se analiza. Puede ser debida a cointervenciones que se dan diferentes en los grupos. Para proteger contra las diferencias en la atención y contra el efecto placebo, puede «cegarse» a aquellos que dan y reciben la atención para que no conozcan el grupo al que están asignados quienes la reciben. Hay resultados que indican que tales sistemas de «cegamiento» son muy importantes en la prevención del sesgos.

Algunos estudios han mostrado que la contaminación (ofrecer la intervención al grupo control) y la cointervención (provisión de atención adicional no intencional a cualquiera de los grupos de comparación) pueden afectar los resultados del estudio. Además, hay evidencia de que los participantes que son conscientes de su grupo de asignación informan más síntomas, y por lo tanto, se obtienen resultados sesgados.

Por estas razones, se debe considerar el uso del «cegamiento» como un criterio de validez. Esto puede hacerse con las siguientes preguntas: ¿Quiénes recibieron atención, desconocían el tratamiento que les había sido asignado? ¿Aquellos que proporcionaron la atención, desconocían el tratamiento asignado?

A menudo se agrega una tercera pregunta relacionada con el «cegamiento»: ¿Los responsables de la evaluación de resultados, desconocían el tratamiento asignado? Esto también tiene relación con el sesgo de detección.

La utilización de la técnica de «cegamiento» siempre es muy importante, pero lo es más en los

estudios que utilizan medidas subjetivas de resultado como el dolor. Es importante desarrollar criterios específicos para valorar la idoneidad del método de «cegamiento» que se usó. En algunas áreas puede ser aconsejable usar el mismo criterio en diferentes revisiones.

Sesgo de desgaste

- El sesgo de desgaste (*attrition bias*) se refiere a las diferencias sistemáticas entre los grupos en la pérdida de participantes durante el estudio.

También se ha denominado "sesgo de exclusión" pero en este artículo lo denominamos sesgo de desgaste o de pérdida para evitar la confusión con los «criterios de exclusión y de inclusión» de los participantes. Cuando la información sobre cómo se tratan las pérdidas (abandonos, bajas o desviaciones de protocolos) de los participantes es insuficiente, los revisores deben tener precaución con los cálculos de seguimiento.

La manera de tratar las pérdidas puede sesgar los resultados y la información incompleta sobre cómo se tratan las pérdidas pueden ocultar este problema.

Sesgo de detección

- El sesgo de detección se refiere a las diferencias sistemáticas en la evaluación de resultados.

Los ensayos que ciegan a quienes evalúan resultados son lógicamente, menos propensos a estar sesgados que en los que no lo hacen.

También, se deberían considerar como indicadores de validez la especificación *a priori* de las medidas de resultado y los análisis que hicieron los investigadores.

MÉTODOS PARA RESUMIR LA VALIDEZ DE LOS ESTUDIOS

Métodos sencillos

Hay varias maneras de evaluar la validez. Una consiste en asignar los valores «se cumple», «no se cumple», o «es poco claro» a los criterios in-

dividuales y al mismo tiempo, evaluar la utilización del «cegamiento» de las asignaciones a los grupos de estudio.

Sin embargo, una vez se han usado varios criterios para evaluar la validez, es aconsejable resumirlos para obtener una evaluación general de los resultados de cada estudio. Un enfoque sencillo para hacerlo es usar tres categorías:

Cuadro N° 2. Evaluación de la Validez

Escala	Interpretación	Relación con los criterios individuales
A Riesgo bajo de sesgo	Sesgo posible, pero es improbable que afecte seriamente los resultados	Se cumplen todos los criterios
B Riesgo moderado de sesgo	Sesgo posible que genera dudas sobre los resultados	Al menos un criterio se cumple sólo parcialmente
C Riesgo elevado de sesgo	Sesgo posible que compromete la confianza de los resultados	Al menos un criterio no se cumple

Las relaciones sugeridas en la tabla serán apropiadas si sólo se usa un número reducido de criterios de evaluación y si todos los criterios tratan de aspectos importantes para la validez de los resultados. De ser posible, los revisores deben obtener información adicional de los autores cuando no queda claro si se cumplió un criterio o no.

Sesgo en los estudios no experimentales

En los estudios observacionales, como los de casos y controles y los de cohortes, es posible controlar y ajustar por factores de confusión conocidos y medibles. Sin embargo, no es posible ajustar por variables que no se sabe si son factores

de confusión o por factores de confusión que no se midieron.

Rara vez puede darse por sentado, que se conozcan todos los factores relacionados con el pronóstico y la respuesta al tratamiento, y en caso de conocerlos pueden surgir dificultades a la hora de medirlos o tenerlos en cuenta en el análisis.

El sesgo de selección puede distorsionar los efectos en cualquier dirección, haciendo que parezcan más grandes o más pequeños de lo que son. Generalmente no es posible predecir la magnitud y a menudo ni siquiera la dirección de este sesgo en estudios específicos. Sin embargo, por término medio, tiende a hacer que los efectos del tratamiento parezcan más grandes de lo que son.

A pesar de estas inquietudes, a veces hay fundamentos para confiar en los estudios observacionales para obtener información acerca de los efectos de las intervenciones de la atención e incluir estos estudios en las RS. Por ejemplo, estudios observacionales bien diseñados han proporcionado información útil sobre los efectos de las intervenciones como el uso obligatorio de los cascos por los motociclistas, la tamización para cáncer de cuello uterino, la difusión de guías de práctica clínica para cambiar la práctica profesional y los efectos adversos poco frecuentes de una medicación.

Se han sugerido diversos criterios a fin de evaluar críticamente la validez de los estudios observacionales. En general, los mismos cuatro tipos de sesgo comentados anteriormente pueden aplicarse a otros tipos de estudios comparativos, como se ilustra a continuación:

Cuadro Nº 3. Sesgos en Estudios Observacionales

Tipo de sesgo	Estudios de cohortes	Estudios de casos y controles
Sesgo de selección	Control de factores de confusión (<i>confounders</i>)	Apareamiento
Sesgo de realización	Medida de la exposición	Medida de la exposición
Sesgo de desgaste	Seguimiento completo	Seguimiento completo
Sesgo de detección	«cegamiento»	Definición de caso y «cegamiento»

Los aspectos relacionados con el sesgo de desgaste son similares en el caso de los ensayos clínicos, los estudios de cohortes y los estudios de casos y controles, y en cuanto a su alcance, se relacionan con el grado en que los individuos incluidos en un estudio son tenidos en cuenta adecuadamente en los resultados.

En lo relacionado con el sesgo de detección también son similares en los estudios de cohortes y van asociados a la definición de caso que se usa en los estudios de casos y controles, ya que los participantes se incluyen en tales estudios teniendo en cuenta el resultado de interés.

La diferencia principal entre los ensayos clínicos y los estudios observacionales guarda relación con el sesgo de selección y con la necesidad de identificar y tener en cuenta los posibles factores de confusión en los estudios observacionales.

Para ello, se deben valorar qué factores de confusión son importantes y en qué medida se midieron y controlaron adecuadamente. La evaluación del sesgo de realización también es más difícil en los estudios observacionales ya que es necesario medir la exposición a la intervención de interés y asegurar que no haya diferencias en la exposición a otros factores que podrían afectar los resultados.

Es importante evaluar si la exposición se midió de manera similar y no sesgada en los grupos que se comparaban.

En resumen, es necesario conocer bien el tema al evaluar la validez de los estudios obser-

vacionales. También se necesita experiencia cuando se evalúa la validez de los ensayos clínicos, pero la propia naturaleza de los estudios observacionales los hace más difíciles de evaluar críticamente. Esta evaluación requiere una comprensión detallada tanto del problema objeto de la revisión como de los aspectos metodológicos.

APLICACIÓN DE LOS CRITERIOS DE EVALUACIÓN CRÍTICA

- Se debe tener en cuenta la formación de los revisores y si éstos deben tener experiencia previa en diseño de estudios y en evaluación crítica de los mismos.

Cuando se decide tener múltiples revisores, debe haber un procedimiento o regla de decisión explícita *a priori* para identificar y resolver los desacuerdos. Independientemente del número de revisores, para asegurar que los criterios de evaluación pueden aplicarse consistentemente, es importante probar cualquier criterio de evaluación planteado en una muestra piloto de artículos.

¿Debe formarse específicamente a los revisores en métodos de investigación, en el área temática de una revisión o en ambos?

Aunque los expertos en las áreas temáticas pueden tener opiniones preconcebidas que llegaran a sesgar sus evaluaciones, es posible que evalúen la validez de los ensayos de manera más consistente que las personas sin experiencia sobre el

tema. Además pueden hacer valiosas apreciaciones distintas a las que podrían hacer personas que sólo tienen experiencia metodológica.

- Sería aconsejable usar tanto expertos en el tema como no expertos y asegurar que todos tienen una comprensión adecuada de los aspectos metodológicos pertinentes.

Incorporación de las evaluaciones de la validez de los estudios en las revisiones

Hay distintas formas de utilizar las evaluaciones de la validez en una revisión:

- Como un umbral para la inclusión de estudios
- Como una posible explicación de las diferencias en los resultados entre los estudios en los análisis de sensibilidad
- Como pesos en el análisis estadístico de los resultados

En algunas revisiones, si el riesgo de sesgo es elevado, se justifica la exclusión de los estudios que no cumplen uno o varios criterios de validez.

Por ejemplo, para medidas de resultado muy subjetivas como el dolor, los revisores pueden decidir incluir sólo los ensayos que previenen el sesgo mediante el «cegamiento» de los participantes.

La decisión sobre dónde situar el punto de corte para la inclusión se puede conceptualizar como una línea continua que va desde "libre de sesgo" a "sin duda sesgado", tal como se ilustra a continuación:

Diagrama N° 1. Conceptualización de Sesgo



Si los revisores elevan el punto de corte metodológico para incluir estudios, se reduce la variabilidad en la validez de los estudios incluidos. En este caso, las evaluaciones de validez situarían los estudios, según el riesgo de sesgo, en el intervalo por encima del punto de corte. Con un punto de corte suficientemente elevado, la variabilidad en la validez de los estudios incluidos es muy pequeña.

RECOLECCIÓN Y ANÁLISIS DE DATOS

El *software* esencial para las Revisiones Cochrane es el programa RevMan. Este *software* facilita el manejo de datos, los análisis cuantitativos y la generación de informes. También se cuenta con otros programas, como:

Componentes claves de un formulario de recolección

Al adaptar o diseñar un formulario de recolección, se deben considerar en primer lugar la cantidad de información que propone recoger. Una

recolección de información demasiado detallada puede dar lugar a formularios más largos que los del estudio original, tediosos y que requieren un tiempo excesivo para los revisores. Por otro lado, si los formularios no se detallan suficientemente y se omiten datos claves, puede ocurrir que se deba usar formularios suplementarios o se tenga que revisar de nuevo un estudio lo cual es lento y frustrante.

Identificación de los estudios

Debe asignársele una identificación a cada estudio que se introducirá en el programa (RevMan u otro seleccionado) cuando se inicie la digitación de la información acerca de las referencias; debe usarse para todos los informes que se refieran al mismo estudio. La identificación se usa en las tablas de estudios incluidos y excluidos y en las tablas de datos de RevMan.

Los revisores pueden decidir si usan el apellido del primer autor y el año de publicación, el país o ubicación dónde se realizó un estudio u otros identificadores informativos. Los estudios pueden ordenarse alfabéticamente según estos códigos.

Es buena idea registrar el origen de la información relevante, si esta información se encontró en un informe o si se obtuvo de fuentes inéditas o de comunicaciones personales. Cualquier información inédita que se use debe estar escrita y debe codificarse de la misma manera que la información publicada.

Verificación de la elegibilidad de un estudio

Aunque la búsqueda y el proceso de selección deberían haber descartado la mayoría de estudios no elegibles a través de la aplicación de los criterios de inclusión, es bueno comprobar los criterios para ser elegibles en el momento del resumen o de la recolección de datos.

La verificación debe hacerse tempranamente ya que el resto del formulario se refiere a los estudios que reúnen los criterios de inclusión, si no se hace así, esto significa malgastar recursos.

Las RS incluyen una tabla de estudios excluidos. En el formulario de recolección puede incluirse un código para identificar las razones por las que se excluyeron estudios. Por ejemplo, supongamos que sólo se quiere incluir en la revisión los ensayos completamente aleatorizados. Una pregunta de verificación en el formulario de recolección puede ser *¿Estudio aleatorizado? Sí, No, No se sabe*. Si un estudio usara asignaciones alternativas, la respuesta a la pregunta es No, y esta información se introduciría como motivo de exclusión.

Características de los estudios

Para evaluar críticamente los estudios hay que analizar y asignar códigos a algunas de sus características específicas. Estas características

pueden agruparse tal como se estructura la información que se entrará en el programa RevMan:

- Métodos
- Participantes
- Tipo de intervenciones
- Indicadores de medidas de resultado.
- La información sobre los participantes suele incluir los datos relacionados con el ámbito de estudio y los criterios diagnósticos para el problema de salud de interés.

Métodos

Los diferentes métodos de investigación pueden influir en las medidas de resultado de los estudios, introduciendo sesgos.

Otras características importantes del método son la duración del estudio; el tipo de ensayo (diseños paralelos, cruzados); el «cegamiento» de pacientes, proveedores y evaluadores de resultados; el número de pérdidas y de cambios de tratamiento; las cointervenciones y otros factores de confusión potenciales.

- El apartado de métodos del formulario de recolección de datos debe incluir los criterios de validez (control de sesgos) utilizados.

Participantes

Las características de los participantes pueden variar sustancialmente entre estudios. Edad y sexo son características útiles para evaluar la aplicabilidad. Otras variables sociodemográficas como el nivel de educación o las variables que tratan la presencia de comorbilidades también son importantes.

A veces los estudios se realizan en regiones geográficas con diferencias culturales importantes

que podrían afectar a la intervención y sus resultados. También las diferencias temporales indican diferencias tecnológicas importantes. Si tales elementos son importantes para la interpretación de la revisión, deben evaluarse.

Los criterios diagnósticos utilizados para definir el problema de salud de interés son una fuente importante de variabilidad clínica y deben describirse.

Intervenciones

Debe describirse la intervención. Más si es compleja como en los estudios que evalúan las psicoterapias, los enfoques educativos o de conducta o las estrategias de prestación sanitaria. Es importante recoger información acerca de quién llevó a cabo la intervención, su contenido, formato, tiempo, etc.

Medidas de resultado de los estudios

Esta puede ser una de las secciones más difíciles de diseñar del formulario de extracción de datos. Los investigadores de los estudios primarios suelen presentar más de una medida de resultado (mortalidad, morbilidad, calidad de vida, etcétera); pueden presentar la misma medida de resultado usando diferentes escalas de medición, pueden incluir medidas de resultado para distintos subgrupos de pacientes o pueden presentar medidas de resultado en diversos momentos del tiempo.

Formato de codificación e instrucciones para codificadores

Es sumamente importante utilizar un sistema de codificación correcto. Los formatos de codifica-

ción utilizados no deben ser complicados para evitar que la persona que extrae los datos se confunda o tome decisiones equivocadas. Los revisores deben disponer de instrucciones y reglas de decisión en el formulario de recolección de datos.

Confiabilidad

La confiabilidad se refiere al grado en que diferentes personas coinciden en la revisión de un estudio. Cuando más de una persona examina los datos, inevitablemente habrá desacuerdos. Los revisores múltiples deben planificar cómo compararán la información de sus formularios de recolección y cómo alcanzarán el consenso en caso de desacuerdo.

El consenso puede lograrse mediante discusión entre revisores o utilizando un árbitro adicional independiente.

ANÁLISIS Y PRESENTACIÓN DE RESULTADOS

¿Por qué usar la estadística en una revisión?

Para preparar una revisión los investigadores deben recoger datos de los estudios individuales, de la misma forma que se recogen datos de los pacientes en los estudios primarios. En ambos casos, pueden usarse los métodos estadísticos para analizar y resumir los datos. Si se usan apropiadamente, proporcionan una herramienta potente para llegar a unas conclusiones que tengan sentido y ayudan a evitar errores de interpretación.

Cuando en las revisiones no se usan métodos estadísticos, es fácil cometer el error de comparar el número de estudios «positivos» con el número de estudios «negativos». Con este enfoque de «recuento de votos», un estudio puede considerarse «positivo» en una revisión y «negativo» en otra, según cómo interpreten los resultados los revisores. Al contar los votos también se tiende a pasar por alto los efectos pequeños pero clínicamente importantes, en particular a contar como «negativos» los estudios con resultados "no estadísticamente significativos".

Otro error que puede ocurrir fácilmente cuando no se usan métodos estadísticos es la ponderación inadecuada de los resultados de los estudios individuales. Por ejemplo, el "recuento de votos" da igual peso a los estudios, ignorando las diferencias en el tamaño y la calidad de los estudios. También es posible que el revisor, cuando usa métodos no cuantitativos, introduzca un sesgo al dar inadecuadamente más peso a los resultados de un estudio respecto a otro.

Por supuesto, el uso de métodos estadísticos no garantiza que los resultados de un meta-análisis sean válidos, como ocurre también en un estudio primario. Como cualquier herramienta, los métodos estadísticos pueden utilizarse mal.

Cuándo no se debe usar la estadística en una revisión

Una revisión sistemática que no incluya un análisis estadístico puede tener el mismo valor que un meta-análisis. Lamentablemente, los revisores a veces no lo creen y piensan que han fracasado si no pueden hacer un meta-análisis.

Tal como se comentó las RS principalmente intentan proporcionar resúmenes no sesgados y

actualizados de lo que se conoce y de lo que no se conoce acerca de los efectos de diferentes formas de atención. El objetivo es ayudar de manera práctica a la toma de decisiones. Los resultados del análisis estadístico (meta-análisis) contribuyen a esto de manera importante, pero éste es un elemento más en una revisión, a menudo no el más importante y si ha sido hecho inadecuadamente, puede ser engañoso.

- Hay dos razones importantes para **no** usar la estadística en una revisión. La razón principal para no usar el meta-análisis es la falta de datos válidos y relevantes. Esto debería ser obvio, pero cuando se presta excesiva atención al meta-análisis, los revisores pueden llegar a incluir estudios de relevancia marginal o de calidad deficiente, usar datos cuestionables y sacar conclusiones poco sólidas.

Este argumento se basa en el principio de que disponer de algunos datos, cualquiera que sea su calidad, es mejor que no tener ninguno. No obstante, *disponer de datos erróneos es peor que no tenerlos*. Por otra parte, saber que no hay ningún dato fiable sobre el cual basar una decisión puede ser de ayuda para las personas que toman decisiones. Esto puede llevar a que se realicen estudios para obtener los datos que se necesitan y que permitirán tomar las decisiones que de otro modo no se habrían tomado.

Otra razón importante para no hacer un meta-análisis es que a veces no tiene sentido. Algunos meta-análisis son difíciles de comprender y pueden combinar erróneamente los resultados de los estudios.

En general, no debe realizarse un meta-análisis si no está claro cómo pueden ayudar sus resultados a las personas que toman decisiones.

¿Qué implica el análisis?

Aunque el uso de métodos estadísticos en las revisiones puede ser sumamente útil, el elemento más esencial de un análisis es la aproximación con detenimiento a sus elementos cualitativos y cuantitativos. Esto implica la consideración de las siguientes preguntas:

- ¿Qué comparaciones deben hacerse?
- ¿Qué resultados de los estudios deben usarse en cada comparación?
- ¿Son los resultados de los estudios similares en cada comparación?
- ¿Cuál es la mejor medida resumen del efecto en cada comparación?
- ¿Qué grado de fiabilidad tienen estas medidas de resumen?

El primer paso para responder estas preguntas consiste en decidir qué comparaciones se quieren hacer. El paso siguiente es preparar resúmenes en forma de tablas de las características y los resultados de los estudios incluidos en cada comparación.

Se puede continuar analizando las diferencias entre los estudios de una manera sistemática, obtener estimaciones del efecto de los estudios y valorar la confianza que debe depositarse en aquellas estimaciones.

¿Qué comparaciones se deben hacer?

El primer paso y el más importante en la planificación del análisis consiste en especificar qué comparaciones se harán. Éstas deben relacionarse clara y directamente con las preguntas que se plantean cuando se formula el objetivo de la revisión.

Las comparaciones principales deben especificarse en el protocolo de la revisión. Sin embargo, con frecuencia es necesario variar las comparaciones y agregar otras nuevas a medida que se recogen los datos. Por ejemplo, puede ocurrir que antes de recoger los datos no se sepa qué medidas de resultado podrán analizarse conjuntamente, o que variaciones importantes en la intervención sólo se puedan observar después de recoger los datos.

¿Qué datos se necesitan para cada comparación?

Una vez se han especificado las comparaciones, el próximo paso en el análisis consiste en preparar resúmenes de los datos en forma de tablas.

En general, hay cinco tipos de información útiles para resumir los resultados y facilitar su análisis:

- Variables clave que caracterizan:
 - A los participantes.
 - Las intervenciones.
 - Las medidas de resultado.
- Resultados en unidades naturales que se comprendan fácilmente.
- Resultados estandarizados, si pueden presentarse de forma que faciliten las comparaciones entre estudios y no creen confusión.

Analizar la heterogeneidad y decidir los análisis de los subgrupos

Una vez se ha decidido qué comparaciones hacer y qué datos se necesitan para cada comparación, la próxima pregunta es si los resultados de los estudios son similares dentro de cada comparación. En otras palabras: las diferencias entre

los resultados de los estudios, ¿son mayores que lo que podría esperarse por azar? Para responder a esta pregunta es útil representar gráficamente los resultados.

- Si los intervalos de confianza de los resultados de cada estudio, no se superponen, esto indica que las diferencias son probablemente «estadísticamente significativas». Cuando hay heterogeneidad «estadísticamente significativa», es probable que las diferencias observadas en los resultados sean causadas por factores diferentes del azar.

Sin embargo, cuando se buscan posibles explicaciones para las diferencias - como diferencias en las intervenciones estudiadas -, hay que ser prudente. Incluso en el caso que un meta-análisis incluyera sólo ensayos controlados, aleatorios y bien diseñados. Dado que generalmente los estudios difieren en muchos aspectos, es arriesgado atribuir las diferencias entre los resultados a alguno de los factores.

Independientemente del grado de heterogeneidad de los resultados, a menudo interesa analizar una categoría específica de participantes en una revisión, como pueden ser, por ejemplo, las mujeres, un cierto grupo de edad o personas con un tipo de enfermedad específico.

Resumen de los efectos de los estudios

El objetivo principal de la mayoría de las RS es proporcionar una estimación confiable de los efectos de una intervención o de un tipo de intervención, basada en una media ponderada de los resultados de todos los estudios de calidad disponibles.

La combinación de los resultados de un grupo de estudios tiene sentido si las diferencias observadas entre los resultados de los mismos no son estadísticamente significativas o prácticamente importantes. En tal caso, la combinación resulta relativamente sencilla.

Cada estudio se resume usando una medida de efecto (*odds-ratio*, riesgo relativo o diferencia de medias) que represente la comparación en el estudio de los grupos de intervención y de control. De esta forma, los participantes en cada estudio se comparan únicamente con las demás personas en el mismo estudio.

A veces se publican meta-análisis en los que se calcula un «tamaño del efecto» para cada grupo de estudio basado en la diferencia de resultados antes y después del tratamiento, y los grupos de diferentes estudios se comparan directamente entre sí. Cuando se hace esto, los resultados de un meta-análisis deben interpretarse con mucho cuidado. Aunque todos los datos provengan de ensayos aleatorios controlados, cuando se utiliza esta metodología la potencia de la aleatorización se pierde por completo y los datos se reducen al equivalente de estudios del tipo «antes-después». Esta aproximación metodológica para la síntesis de los datos debería evitarse.

El intervalo de confianza de una estimación derivada de una síntesis estadística apropiada indica la precisión de la estimación. En principio, no es diferente de un intervalo de confianza de una estimación derivada de un estudio primario. Sin embargo, un aspecto importante que surge en meta-análisis es si incorporar la variabilidad entre estudios cuando se estima el intervalo de confianza (enfoque de «efectos aleatorios»).

Si hay poca variabilidad entre estudios, si el test de homogeneidad tiene un valor p grande, por ejemplo mayor que 0.10), el método utilizado afectará poco al intervalo de confianza. Si hay variabilidad significativa entre estudios, un análisis que la ignore, enfoque de «efectos fijos», dará un intervalo de confianza más estrecho. Si hay heterogeneidad significativa, los revisores deben tener precaución al interpretar la estimación global derivada de un meta-análisis, así como al atribuir las diferencias entre los estudios a cualquier factor.

Cuando hay heterogeneidad sustancial en los resultados hay desacuerdo en la forma de resumirlos y en si estos deben combinarse. Sin embargo, existe un mínimo acuerdo:

- Primero y principal, los revisores deben alertar a los lectores de su revisión cuando haya heterogeneidad sustancial, y deben recomendar una interpretación cautelosa de los resultados agregados.
- Los revisores deben intentar explorar y explicar los motivos de la heterogeneidad, pero las interpretaciones deben ser cautelosas porque los análisis generalmente se hacen *a posteriori*. A menudo, los análisis de subgrupos son útiles pero de difícil interpretación.

Una posible explicación de la heterogeneidad, que puede investigarse *a priori*, es la calidad metodológica. Distinta calidad metodológica entre los estudios puede producir a la vez sesgo y heterogeneidad. Si hay diferencias importantes en la calidad de los estudios, los revisores deben valorar la posibilidad de excluir los estudios de menor calidad.

Cuando hay heterogeneidad manifiesta, esto conduce al terreno de la discusión de los diferentes

métodos de cálculo del intervalo de confianza de una medida de efecto cuando hay heterogeneidad manifiesta. Si la prueba de heterogeneidad es estadísticamente significativa y las diferencias en los resultados son importantes, entonces:

- Los revisores deben considerar las posibles fuentes de heterogeneidad (diferencias en las dosis, pautas o duración del tratamiento, características de los participantes). En particular, los revisores deben explorar si las diferencias en el control del sesgo en los estudios explican la heterogeneidad, por ejemplo, el «cegamiento» de la aleatorización.

Si no puede explicarse la heterogeneidad, los revisores deben considerar las siguientes opciones:

- No agregar los estudios.
- Usar los modelos de efectos fijos interpretando los resultados con cautela.
- Usar los modelos de efectos aleatorios con idénticas prevenciones.
- Usar ambos modelos (efectos fijos y efectos aleatorios) como una ayuda adicional para explicar la incertidumbre que conlleva el análisis de estudios heterogéneos.
- Sea cual sea la decisión que se tome, los revisores deben explicar claramente qué se hizo y por qué.

Presentación y descripción de los análisis

Los métodos analíticos que se usan en las revisiones deben describirse en la sección de Métodos. Los resultados principales deben describirse y explicarse en el texto. Los revisores deben ser cuidadosos en asegurar que el enfoque en la sección de Resultados sea compatible con los objetivos y las comparaciones específicas del

protocolo de la revisión. Los análisis *a posteriori* deberían informarse como tales.

Evaluación crítica: Componentes y características de una buena revisión sistemática

La RS de cualquier tópico de medicina debe realizarse de acuerdo a las mismas recomendaciones de rigurosidad que las utilizadas en estudio originales.

Como ya se mencionó las RS se llevan a cabo como cualquier investigación original pero en lugar de tener individuos y comunidades de observación, se utilizan los resultados de los estudios primarios. Por lo anterior se han establecido una serie de recomendaciones que se deben tener en cuenta al leer y acoger las recomendaciones de las RS.

Es importante que quienes accedan a una RS como herramienta para valorar una situación clínica en particular o para establecer criterios de decisión sobre alguna conducta a seguir en la atención en salud, realicen una valoración crítica que les permita establecer su validez, su metodología y su aplicabilidad.

Entender la evaluación crítica de una revisión sistemática le permitirá determinar si es válida, interpretar los resultados y evaluar su aplicabilidad en la práctica clínica, en la salud pública y/o en el desarrollo de futuras investigaciones.

Al igual que en un estudio de investigación pueden presentarse errores, por sesgo o por azar, que hacen que los resultados finales no describan exactamente la situación estudiada. En las RS donde se evalúan simultáneamente los hallazgos de un conjunto de investigaciones también

pueden darse situaciones que disminuyan su calidad o pongan en duda la confiabilidad de sus análisis o sus conclusiones.

Para la apreciación crítica de una RS se presentan las siguientes guías basadas en preguntas, contenidas en diferentes etapas de evaluación, que los lectores deben resolver con respecto al contenido, metodología y resultados de una RS.

Recomendaciones para lectura crítica de RS del CASP*

El CASP (*Critical Appraisal Skills Programme*) programa de habilidades en lectura crítica para ayudar a los «decisiones» del Servicio de Salud Inglés a desarrollar habilidades para leer críticamente sobre la evidencia científica de la eficacia clínica. Trabaja con programas locales para promover cuidados de salud basados en la evidencia y colabora con el Centre for Evidence-Based Medicine (Centro de la Medicina Basada en la Evidencia) de la Universidad de Oxford que enseña a los clínicos cómo tomar decisiones, basadas en la evidencia, sobre un paciente concreto. (http://www.hrc.es/caspe/preguntas_rev.html)

Preguntas para ayudarte a hacer lectura crítica

Hay tres aspectos generales a tener en cuenta cuando se hace la lectura crítica de un artículo:

¿Son válidos sus resultados?

¿Cuáles son los resultados?

¿Son aplicables en mi medio?

* Tomado y adaptado del material electrónico publicado en <http://populi.eest.ucr.ac.cr/cursoweb/4213evcrit.htm>

Las preguntas están diseñadas para ayudarte a pensar sistemáticamente sobre estos aspectos. Contienen unas primeras preguntas «de eliminación» que se pueden responder rápidamente. Sólo si la respuesta es «sí» en todas ellas vale la pena continuar con las preguntas restantes.

• **Evaluación de la validez interna**

El primer paso en la apreciación crítica de una revisión es establecer la calidad metodológica de dicho estudio para determinar la confianza que el lector puede tener en sus resultados. En el caso de que se concluya que la revisión no se ha realizado con el rigor metodológico esperado, es poco probable que los resultados reflejen la verdad y por lo tanto no debería leerse, o dicha lectura debería realizarse teniendo en cuenta la forma en que los errores cometidos en su desarrollo pueden distorsionar los resultados.

La RS responde claramente a un tema en cuanto a:

¿Población de pacientes?

¿Intervención que se desea evaluar?

¿Resultados que se desean medir?

¿Están especificados los criterios de selección del tipo de investigaciones que se incluirán en la revisión?

De acuerdo a la pregunta de la RS.

El tipo de diseño de investigación que se va a seleccionar

Está aclarado el método que se utilizó para la búsqueda de las investigaciones?

¿Cuál es la probabilidad de que se hayan omitido investigaciones relevantes?

¿Se hizo la evaluación para valorar la calidad de los estudios incluidos?

¿La evaluación de calidad de los estudios fue realizada por más de una persona en forma independiente estableciéndose el acuerdo que hubo entre ellos?

• **Interpretación de los resultados de la revisión**

Una vez que se ha evaluado la calidad metodológica de la revisión y llega a la conclusión de que se justifica la lectura de sus resultados se deben interpretar correctamente los mismos según las siguientes técnicas para la interpretación cuantitativa de los resultados.

¿Son consistentes los resultados de estudio a estudio?

¿Cuáles fueron los resultados generales de la revisión?

¿Cuán precisos fueron?

• **Evaluación de la validez externa**

Luego de haber interpretado los resultados de la revisión se deberá evaluar si los mismos son aplicables a «nuestros pacientes», y si los beneficios superan a los daños potenciales.

¿Mis pacientes son similares a los incluidos en los estudios originales?

¿La intervención es factible de aplicar en mi medio?

¿Se consideraron todos los resultados clínicamente importantes?

¿Superan los beneficios a los daños potenciales?

Preguntas basadas en el formulario para revisores de artículos de RS del Canadian Association Journal (Jenicek M. 1996)

Resumen

¿Se especifica el propósito de la revisión?
¿Están los métodos de investigación claramente descritos?
¿Están los hallazgos importantes claramente resumidos?
¿Están las conclusiones y recomendaciones mas importantes claramente delineadas?

Introducción

¿Está el propósito de la revisión claramente expresado?
¿Están la estrategia de búsqueda y la identificación de las fuentes relevantes claramente descritas?
¿Están los criterios de inclusión y exclusión claramente identificados?

Hallazgos

¿Fue valorado objetivamente la validez de los artículos incluidos?
¿Fueron identificadas las limitaciones de los resultados de los artículos relevantes?
¿Fue analizada críticamente la variación de los resultados de los estudios pertinentes?
¿Fueron combinados apropiadamente los resultados de los estudios pertinentes?
(En caso de haber sido aplicado el metaanálisis).

Conclusiones

¿Se proporciona un resumen claro de los hallazgos pertinentes?

¿Están apoyadas las conclusiones de los revisores por la evidencia proporcionada?
¿Se proponen directivas específicas para nuevas investigaciones?

BIBLIOGRAFÍA

- Annals of Internal Medicine EDITORIALS. Systematic Review: Critical links in the Great Chain of Evidence. Annals of Internal Medicine. 1 March 1997. 126: 389 - 391.
- Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. J R Statist Soc A 1988; 151: 445-63.
- Cooper HM, Ribble RG. Influences on the outcome of literature searches for integrative research reviews. Knowledge 1989; 10: 179-201.
- Cooper H, Hedges LV, eds. The handbook of research synthesis. Nueva York: Russell Sage Foundation, 1994.
- Chalmers I, Altman DG, eds. Systematic reviews. Londres: Br Med J Pub. Group, 1995.
- Cooper H, Hedges LV, eds. The handbook of research synthesis. New York: Russell Sage Foundation, 1994.
- Chalmers I. Applying overviews and meta-analysis at the bedside: Discussion. J Clin Epidemiol 1995, 48 (1):67-70.
- Dickersin K, Min YI. NIH clinical trials and publication bias. Online J Curr Clin Trials [serial online] 1993; Doc No 50.
- Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. BMJ 1994; 309:1286-91.

- Freemantle N, Bero L, Grilli R, Grimshaw J, Oxman A (eds.) Effective Professional Practice Module of The Cochrane Database of Systematic Reviews. Available in: The Cochrane Library [database on disk and CD]. The Cochrane Collaboration; Issue 3. Oxford: Update Software, 1996. Publicación actualizada trimestralmente.
- Glenton C, Oxman AD. Cochrane Review Methodology Database. In: The Cochrane Library [database on disk and CD]. The Cochrane Collaboration. Oxford: Update Software; 1996-. Publicación actualizada trimestralmente.
- Jackson GB. Methods for integrative reviews. *Rev Educ Res* 1980; 50: 438-60.
- Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clin Trials* 1996; 17: 1-12.
- Jenicek M. Epidemiología: La lógica de la medicina moderna. Masson, S.A. Barcelona - Madrid - París - Milán. 1996.
- Kunz RA, Oxman AD. Empirical evidence of selection bias in studies of the effects of health care: a systematic review. Presented at the Cochrane Colloquium, Oslo, 5-8 October, 1995.
- Light RJ, Smith PV. Accumulating evidence: procedures for resolving contradictions among different research studies. *Harv Educ Rev* 1971; 41: 429-71.
- Manual de la Colaboración Cochrane (versión española de "The Cochrane Collaboration Handbook") [actualización de Septiembre 1997]. Sabadell: Centro Cochrane Español; 1998.
- Marti J, Bonfill X, Mejía R, Tohá D. Por una atención pediátrica basada en la evidencia científica. La Colaboración Cochrane. *Revista Pediatría de Atención Primaria*. Vol. I: 87 a 100. Número 1. Enero - marzo 1999.
- Naylor CD. The case for failed meta-analyses. *J Evaluation Clin Practice* 1995; 1: 127-30.
- Ortiz Z. Meta-análisis como método básico en la investigación científica y en la práctica diaria. *Boletín de la Academia Nacional de Medicina*. Buenos Aires. 1997, 75 (2do semestre): 469 - 75.
- Oxman AD, Guyatt GH et al, Preguntas para ayudar a entender una revisión sistemática Oxman AD, Guyatt GH et al, *Users' Guides to The Medical Literature*, VI How to use an overview. (*JAMA* 1994; 272 (17): 1367-1371)
- Oxman AD, ed. Section VI: preparing and maintaining systematic reviews: The Cochrane Collaboration handbook. Oxford: Cochrane Collaboration, 1994.
- Petitti DB. Meta-analysis, decision analysis and cost-effectiveness analysis: methods for quantitative synthesis in medicine. Nueva York: Oxford University Press, 1994.
- Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273:408-12.
- Thacker SB. Meta-analysis: a quantitative approach to research integration. *JAMA* 1988; 259: 1685-9.
- Wilson A, Henry DA. Meta-analyses. Part 2: Assessing the quality of published meta-analyses. *Med J Austr*. 1992; 156: 173-87.