Manual de revisores 4.1

Actualización Junio de 2000

© La Colaboración Cochrane, 2000.

Contenido

ACERCA DEL *HANDBOOK*

Editores

Andrew Oxman y Mike Clarke

Cómo se debe citar esta versión del *Handbook*

Clarke M, Oxman AD, editores. Cochrane Reviewers' Handbook 4.1 [updated June 2000]. En: Review Manager (*RevMan*) [Computer program]. Versión 4.1. Oxford, England: The Cochrane Collaboration, 2000.

Cuando se refiera a una sección o sub-sección específica, refiérase a ello por el título y el número de la sección, NO por los números de páginas; por ejemplo:

Clarke M, Oxman AD, editores. Formulating the problem. Cochrane Reviewers' Handbook 4.1 [updated June 2000]; Section 4. In: Review Manager (*RevMan*) [Computer program]. Version 4.1. Oxford, England: The Cochrane Collaboration, 2000.

Cómo citar el Manual de Revisores (versión castellana del *Cochrane Reviewers' Handbook)*

Manual de Revisores (versión española de "*Cochrane Reviewers' Handbook*") [actualización de Junio 2000]. Barcelona: Centro Cochrane Iberoamericano; 2001.

Direcciones de contacto:

Dr Andy Oxman
Health Services Research Unit
National Institute of Public Health
P.O. Box 4404 Torshov
0462 Oslo
NORWAY
Tel: +47 22 04 26 75
Fax: +47 22 04 25 95
Email: andrew.oxman@labmed.uio.no

Mike Clarke
UK Cochrane Centre
Summertown Pavilion
Middle Way
Oxford OX2 7LG
United Kingdom
Phone: +44 –1865 516300
Fax: +44 –1865 516311
email: mclarke@cochrane.co.uk

Direcciones en Internet:

Direcciones WWW

Alemania
http://www.cochrane.de

Reino Unido
http://www.update-software.com/ccweb/

Dinamarca
http://www.cochrane.dk/default.html

España
http://www.cohrane.es/

Australia
http://som.flinders.edu.au/fusa/cochrane

Canadá
http://hiru.mcmaster.ca/cochrane

Japón
http://www.nihs.go.jp/acc/default.html

USA
http://www.cochrane.org

Servidores FTP

Australasia:
Conectar con: healthinfo.fmc.flinders.ed.au
login: cochrane
password: australia (must be in lower case)

Reino Unido**:**
Conectar con: ftp.cochrane.co.uk
login: anonymous
password: <your e-mail address>

Canadá**:**
Conectar con: hiru.mcmaster.ca
login: cochrane
password: archie

Fuentes de apoyo

Fuentes de apoyo actuales

National Institute of Public Health (Norway)

Fuentes de apoyo anteriores

Health Services Research and Development Service and the University of Texas Health Science Center, San Antonio (USA)
US Veterans Health Administration (USA)
National Health Service Research & Development Programme, England
Oxford Regional Health Authority, UK
Nuffield Provincial Hospitals Trust, UK
LW Frohlich Fund, USA.
Norwegian Ministry of Health and Social Affairs (Norway)
Norwegian Research Council (Norway)
Glaxo Wellcome (Norway)
Pfizer (Norway)

Agradecimientos

Muchas personas han contribuido a esta versión y a las precedentes del Manual de Revisores (the Cochrane Reviewers' Handbook) de la Colaboración Cochrane, tanto directa como indirectamente (a través de talleres, sugerencias útiles y buenas ideas). Vaya este agradecido reconocimiento a todos los que han hecho sus valiosas aportaciones al acoger con entusiasmo la Colaboración Cochrane y comprometer su propio tiempo para hacer de todo ello una realidad.

Los editores del Manual son Mike Clarke y Andy Oxman, quienes reciben asesoría del Grupo Asesor del Manual (*Handbook Advisory Group)*. Además de los editores, los miembros actuales de este Grupo son:
Chris Cates, Jon Deeks, Matthias Egger, Alex Jadad, Philippa Middleton, Ole Olsen, y los representantes de los Grupos Cochrane de Métodos (*Cochrane Methods Groups*).
También queremos agradecer a las siguientes personas por sus constantes contribuciones a la presente edición del Manual: Doug Altman, Hilda Bastian, Lisa Bero, Monica Fischer, Davina Ghersi, Claire Glenton, Jeremy Grimshaw, Jeph Herrin, Andrew Herxheimer, Carol Lefebvre, David Moher, Rasmus Moustgaard, Melissa Ober, Drummond Rennie, Jacob Riis, Ken Schulz, Lesley Stewart, Luke Vale, Veronica Yank.

La primera edición del Manual (1994) fue desarrolada por Andy Oxman, Iain Chalmers, Mike Clarke, Murray Enkin, Ken Schulz, Mark Starr, Kay Dickersin, Andrew Herxheimer y Chris Silagy con el apoyo administrativo de Sally Hunt.

La segunda y tercera edición del Manual (1995 and 1997) fueron editadas por Andy Oxman y Cynthia Mulrow.

¿QUÉ CONTENIDO ES NUEVO?

Cambios y correciones en la versión 4.1.0 (Junio de 2000) del Manual

Correciones más importantes y cambios

Capítulo 2: se han agregado guías para la publicación de revisiones Cochrane en revistas.
Capítulo 5: ha sido actualizado.
Capítulo 6: ha sido actualizado.
Capítulo 11: se ha añadido una nueva sección (11.6) acerca de la conversión de revisiones que utilizan datos individuales de pacientes al formato de una revisión Cochrane.
Apéndice 2a, sinopsis: La guía para preparar las sinopsis ha cambiado con el objeto de reflejar la nueva política que indica que la responsabilidad de aprobar la sinopsis para que sea incluida en una revisión Cochrane recae únicamente en el grupo de revisión relevante.
Apéndice 2a: se ha añadido una sección para mostrar los elementos de los protocolos y revisiones Cochrane que deben publicarse.

Correcciones menores

Agradecimientos: se reconoce la ayuda de personas adicionales en la preparación de esta versión de Manual.
Apéndice 2a, Texto: La importancia de mantener las búsquedas actualizadas hasta la fecha se ha añadido a las guías sobre el contenido de la sección de Estrategia de búsqueda dentro del texto de una revisión Cochrane.
Apéndice 2a, referencias: el título de la referencia de Flanagin 1998 ha sido corregido.

Correciones y cambios en la Versión 4.0.0 (Julio de 1999) del Manual

El Manual ha sido cuidadosamente revisado para que tenga en cuenta los cambios en *RevMan*. También se han actualizado otras secciones del Manual.

Correcciones y cambios en la Versión 3.0.2 (Septiembre 1997) del *Manual* (Cochrane Library 1997, Issue 4)

Correcciones más importantes y cambios.

1. En el apéndice 2c, “condiciones de publicación”, ahora se ha especificado que debería completarse un nuevo formulario de “condiciones de publicación” con cada nueva versión de una revisión.
2. Con el fin de guardar una consistencia entre los números de versión del Manual con los números de versión del *RevMan*, el primero utilizará a partir de ahora 3 dígitos:

- El primer dígito indica una nueva versión del *RevMan* y del Manual.
- El segundo dígito indica una edición provisional del *RevMan* y del Manual.
- El tercer dígito indica cambios sólo en el Manual.

Correcciones menores y cambios

1. La sección 5.5 sobre búsquedas manuales ha sido actualizada para incorporar el desarrollo del registro de control de un estudio que pueda ser relevante para su inclusión en las revisiones Cochrane (CENTRAL).
2. El glosario tiene cuatro incorporaciones: “CENTRAL", "trend", "Trials Registers Development Group", y "peer review". Los términos "Handsearching" y "Cochrane Controlled Trials Register (CCTR)" han sido actualizados.
3. Synapse Publishing Inc. ha puesto una versión del Manual en la *web* en la siguiente dirección: http://www.medlib.com/cochranehandbook/.
4. Se han hecho correcciones a las referencias del apéndice 2a.6.
5. La lista de las versiones del Manual y de materiales relacionados ha sido actualizada.
6. Se han actualizado las secciones “Acerca del Manual” y “Qué contenido es nuevo”.

Correcciones y cambios en la versión Version 3.0.1 (Diciembre de 1996) del Manual (Cochrane Library 1997, Issue 3)

Correcciones más importantes y cambios

1. Apéndice 11a, 'Metodología práctica del metanálisis (revisión) utilizando datos de pacientes individuales actualizados” ha sido añadido al Manual.
2. Apéndice 5a, 'Registros de ensayos clínicos” fue actualizado.

Correcciones menores y cambios

1. Se actualizaron todas las referencias de las publicaciones incluidas en la Cochrane Library ('Cómo citar el Manual'; referencias: sección 1; referencias: sección 3; referencias: sección 4; referencias: sección 6; referencias: sección 8; Apéndice 5b; Apéndice 5c.).

Correcciones y cambios en la Versión 3.0.0 (Octubre 1996) del Manual (Cochrane Library, Issue 1)

1. Responsabilidad Editorial

Se delegó la responsabilidad de mantener el material anteriormente incluido en las secciones I a V del Manual de Revisores según se describe a continuación. El Manual está constituido ahora únicamente por lo que era anteriormente la Sección VI: Preparación y Mantenimiento de Revisiones Sistemáticas (Oxman, 1995). Cynthia Mulrow, directora del Centro Cochrane de San Antonio (*San Antonio Cochrane Center*) se unió a Andy Oxman como co-editor. Se ha revisado completamente el Manual, en respuesta a las sugerencias que hemos recibido con respecto a la edición anterior del Manual y al Manual de Entrenamiento preparado por el Centro Cochrane de San Antonio (*San Antonio Cochrane Center*).

Las responsabilidades editoriales acerca del material escrito preparado en nombre de la Colaboración Cochrane han ido evolucionando, y en 1995 quedó claro que se requería de nuevos ajustes para enfrentar las nuevas circunstancias. En su reunión

del 27 de Febrero de 1996 en San Francisco, el *Steering Group* estableció un Comité Editorial para que supervisara la preparación del material escrito desarrollado en nombre de la Colaboración. Este es uno de los cinco grupos responsables de funciones centrales que se reportan directamente al *Steering Group*. Los otros grupos encargados de funciones centrales son: El Grupo de Desarrollo de Software (*Software Development Group*), el Grupo de Desarrollo de Registros de Ensayos (*Trials Registers Development Group*), un grupo encargado de organizar los próximos Coloquios, y el equipo editorial del Manual.

Iain Chalmers y Andy Oxman han propuesto cambios adicionales en las responsabilidades editoriales, con el fin de tener en cuenta diversos desarrollos, que incluyen:

- Una potencial duplicación de esfuerzos y de confusión acerca de las funciones del Comité Editorial y del equipo editorial del Manual
- La disponibilidad de la CDSR y el desarrollo de módulos en la CDSR para los Centros Cochrane, los Ambitos, los Grupos de Métodos, la Red de Consumidores y los GCR.
- El establecimiento de un *Steering Group* electo, que tenga representantes de cada entidad y la formación de grupos responsables de cumplir funciones centrales, y que se reporten directamente al *Steering Group.*

Los cambios propuestos circularon entre todos los grupos registrados y fueron aprobados por el *Steering Group* en su reunión del 19 de Agosto de 1996. Las nuevas disposiciones son las siguientes:

Material acerca de	**Grupo responsable**	**Coordinador actual**
La Colaboración	Comité Editorial	Jos Kleijnen
Funciones centrales:		
Manual	Grupo Asesor del Manual	Andy Oxman & Cynthia Mulrow
Software	Grupo de Desarrollo de Software	Monica Fischer
Registros de ensayos	Grupo de Desarrollo de Registros de Ensayos	Kay Dickersin & Jean-Pierre Boissel
Grupos registrados:		
GCR	Representantes de los GCR en el *Steering Group*	A decidir
Centros Cochrane	Directores de Centros en el *Steering Group*	Peter Gø tzsche
Ambitos	Representantes de los ámbitos en el *Steering Group*	A decidir
Red de Consumidores	Representantes de la Red de Consumidores en el *Steering Group*	A decidir
Grupos de Métodos	Representante de los Grupos de Métodos en el *Steering Group*	Andy Oxman

2. Resúmenes

Los resúmenes dejan de ser opcionales y los encabezados en los resúmenes han cambiado de la siguiente manera:

- Objetivos
- Estrategia de búsqueda
- Criterios de selección
- Obtención y análisis de los datos
- Resultados principales

(Ver Sección 2a.2 en el apéndice 2a)

3. Descripciones de los métodos utilizados por los Grupos Colaboradores de Revisión

Todas las revisiones deben mencionar explícitamente cuándo se hicieron las búsquedas por última vez en el registro de ensayos mantenido por el GCR responsable de la revisión para identificar estudios relevantes. La descripción de los métodos empleados para desarrollar y mantener los registros de ensayos de los GCR se incluyen en los módulos de los GCR publicados en la Base de Datos Cochrane de Revisiones Sistemáticas (*Cochrane Database of Systematic Reviews - CDSR*). En el módulo del grupo también deben describirse otros métodos estandarizados empleados por el GCR. Los revisores deben mencionar explícitamente que han utilizado estos métodos, y también cuando han utilizado otros métodos distintos a los estandarizados por el grupo.

4. Revisores de evidencia de tipo no experimental

Algunos GCR, Ambitos y Grupos de Métodos (GM) han comenzado a explorar maneras de incorporar evidencia de tipo no experimental en las revisiones, cuando esto sea apropiado. Estos desarrollos se ven reflejados en el cambio de la terminología del término “ensayos” por el término “estudios” y al agregarse el título de “Tipos de estudios” como un nuevo encabezado en la sección de “Criterios de selección”.

5. Vínculos entre el Manual y otras fuentes de información relacionadas

El Manual está siendo vinculado a varias fuentes de información relacionadas (ver “Acerca del Manual”). Estas incluyen: La *Cochrane Review Methodology Database*, el Manual de Entrenamiento del Centro Cochrane de San Antonio, *Review Manager*, un glosario, una lista de preguntas frecuentes, una biblioteca con ejemplos, una biblioteca de diapositivas, un registro de estudios metodológicos empíricos, revisiones sistemáticas de dichos estudios, y módulos preparados por los GM para ser incluidos en el *CDSR*.

7. Conflicto de interés

Deberá incluirse una declaración de conflicto de intereses en todas las revisiones Cochrane a partir del segundo número de la Cochrane Library de 1997 (ver Sección 2a.2 y Sección 2a.4 en el apéndice 2a).

Cambios propuestos que todavía no se han llevado a cabo

Conclusiones

Dado que los resultados de una revisión pueden interpretarse de diferentes formas según la perspectiva o circunstancias de cada uno, el *Steering Group* decidió en 1994 separar las conclusiones de una revisión del resto de la misma, y adjuntar las

conclusiones o los comentarios de los Ámbitos (*Fields*) y otras entidades que representan diferentes perspectivas (junto con las conclusiones de los revisores) a las revisiones publicadas en la CDSR. No obstante, todavía no se han incorporado las medidas prácticas para la preparación y su posterior actualización. Las conclusiones de los revisores -incluidas las implicaciones para la práctica y para la investigación- se mantienen actualmente como parte del texto de cada revisión.

1. INTRODUCCIÓN

Los profesionales sanitarios, consumidores, investigadores y gestores sanitarios se ven desbordados por una cantidad de información imposible de manejar. Necesitamos revisiones sistemáticas para integrar eficientemente toda la información válida y proporcionar una base para tomar decisiones de manera racional (Mulrow 1994). Las revisiones sistemáticas establecen cuándo los efectos de la atención sanitaria 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 cuándo pueden variar significativamente los efectos. El uso de métodos explícitos y sistemáticos en las revisiones limita el sesgo (errores sistemáticos) y reduce los efectos encontrados por azar, proporcionando así resultados más fiables sobre los cuales sacar conclusiones y tomar decisiones (Antman 1992, Oxman 1993b). El metanálisis - uso de los métodos estadísticos para resumir los resultados de estudios independientes entre sí - puede proporcionar estimaciones más precisas sobre los efectos de la atención sanitaria que aquellas derivadas de los estudios individuales incluidos en una revisión (Oxman 1993a, Sacks 1987, L'Abbe 1987, Thacker 1988).

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 (Light 1971, Glass 1976, Rosenthal 1978, Jackson 1980, Cooper 1982). No fue hasta fines de los años ochenta que diferentes personas alzaron su voz para señalar la deficiente calidad científica de muchos artículos de revisión sobre temas sanitarios (Mulrow 1987, Yusuf 1987, Oxman 1988). A pesar de ello, el reconocimiento de la necesidad de revisiones sistemáticas 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 de estudios empíricos acerca de los métodos empleados en las revisiones (EMSMG 1999), el número de revisiones sistemáticas publicadas en revistas de sanitarias (NHS CRD 1999), y el rápido crecimiento de la Colaboración Cochrane (CDSR 1999).

Este Manual constituye el trabajo de un gran número de personas, incluidos aquellos representados en la Base de Datos Cochrane sobre Metodología de las Revisiones (EMSMG 1999), (*Cochrane Review Methodology Database-CRMD*); las aportaciones de los Grupos de Trabajo Cochrane sobre Metodología (*Cochrane Methods Working Groups-MWG*); con la experiencia práctica y sugerencias de los Grupos Colaboradores de Revisión (*Collaborative Review Groups-CRG*)(CDSR 1999) - que han abordado la ingente tarea de examinar sistemáticamente los efectos de la atención sanitaria dentro de sus propias áreas de interés. También los Centros Cochrane, que proporcionan formación a los revisores, han aportado sus comentarios. Siempre que ha sido posible, las recomendaciones hechas aquí se basan en la evidencia empírica y los consejos proporcionados por los Grupos Cochrane de Métodos.

Nuestro objetivo es ayudar a los revisores a que tomen buenas decisiones acerca de los métodos que utilizan y no dictar criterios arbitrarios. Sin embargo, en aquellas circunstancias en que la Colaboración Cochrane ha sentado una política a seguir por los revisores Cochrane, se menciona con claridad. Las orientaciones proporcionadas aquí están concebidas con el fin de ayudar a los revisores a ser sistemáticos y explícitos (¡no mecanicistas!) acerca de las cuestiones que plantean y cómo generar

respuestas a aquellas cuestiones. En ningún caso estas orientaciones pueden sustituir el sentido común.

La Colaboración Cochrane y el Manual de Revisores se centran particularmente en revisiones sistemáticas de ensayos clínicos controlados con distribución aleatoria (RCTs) (*randomised controlled trials*) porque tienen mayores posibilidades de proporcionar información más fiable que otras fuentes de evidencia sobre los efectos diferenciales de las distintas alternativas de atención en salud (Kunz 1998). Las revisiones sistemáticas de otros tipos de evidencia también pueden ayudar a quienes desean tomar las mejores decisiones sanitarias, en particular sobre aquellas prestaciones en las cuales no se han hecho RCTs. Los principios básicos de revisión de la investigación son los mismos, cualquiera que sea el tipo de evidencia que se someta a examen. Aunque nos centramos principalmente en las revisiones sistemáticas de RCTs, también nos ocupamos de temas específicos para la revisión de otras fuentes de evidencia cuando ello resulta pertinente.
Se está desarrollando una guía más completa para este tipo de revisiones.

Las revisiones Cochrane tienen un formato estándar que describimos en la próxima sección (Sección 2). Los que preparen una revisión deben empezar por desarrollar un Protocolo (Sección 3). Las siguientes siete secciones están organizadas siguiendo los pasos que se requieren para la preparación y actualización de la revisión sistemática:

- Formulación del problema.
- Localización y selección de los estudios.
- Evaluación de la calidad de los estudios
- Obtención de los datos.
- Análisis y presentación de los resultados.
- Interpretación de los resultados.
- Mejora y actualización de las revisiones.

En la última sección consideraremos temas específicos relacionados con el uso de datos de pacientes individuales en las revisiones.

1.1 Referencias

Antman 1992. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA 1992; 268:240-8.

CDSR 1999. Cochrane Database of Systematic Reviews. In: The Cochrane Library, Issue 2, 1999. Oxford: Update Software. Updated quarterly.

Cooper 1982. Cooper HM. Scientific guidelines for conducting integrative research reviews. Rev Educ Res 1982; 52:291-302.

EMSMG 1999. Empirical Methodological Studies Methods Group. Cochrane Review Methodology Database. In The Cochrane Library, Issue 2, 1999. Oxford: Update Software; Updated quarterly.

Glass 1976. Glass GV. Primary, secondary, and meta-analysis of research. Educ Res 1976; 5:3-8.

Jackson 1980. Jackson GB. Methods for integrative reviews. Rev Educ Res 1980; 50:438-60.

Kunz 1998. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ 1998; 317:1185-90

L'Abbe 1987. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Ann Intern Med 1987; 107:224-33.

Light 1971. Light RJ, Smith PV. Accumulating evidence: procedures for resolving contradictions among different research studies. Harv Educ Rev 1971; 41:429-71.

Mulrow 1987. Mulrow CD. The medical review article: state of the science. Ann Intern Med 1987; 106:485-8.

Mulrow 1994. Mulrow CD. Rationale for systematic reviews. BMJ 1994; 309:597-9.

NHS CRD 1996. NHS Centre for Reviews and Dissemination. Database of Abstracts of Reviews of Effectiveness. In: The Cochrane Library, Issue 2, 1999. Oxford: Update Software. Updated quarterly.

Oxman 1988. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. Can Med Assoc J 1988; 138:697-703.

Oxman 1993a. Oxman AD. Meta-statistics: Help or hindrance? ACP J Club 1993; 118:A-13.

Oxman 1993b. Oxman AD, Guyatt GH. The science of reviewing research. Ann NY Acad Sci 1993; 703:125-33.

Rosenthal 1978. Rosenthal R. Combining results of independent studies. Psychol Bull 1978; 85:185-93.

Sacks 1987. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med 1987; 316:450-5.

Thacker 1988. Thacker SB. Meta-analysis: a quantitative approach to research integration. JAMA 1988; 259:1685-9.

Yusuf 1987. Yusuf S, Simon R, Ellenberg S (eds). Proceedings of 'Methodologic issues in overviews of randomized clinical trials'. Stat Med 1987; 6:217-409.

2. FORMATO DE UNA REVISIÓN COCHRANE

El formato de las revisiones Cochrane tiene varios objetivos. Ayuda a quienes las consultan a encontrar rápidamente los resultados de la investigación y a evaluar su validez, la aplicabilidad de sus resultados, y sus implicaciones. Guía a los revisores en la presentación de su trabajo, explícitamente y con concisión, y reduce al mínimo el esfuerzo requerido para hacerlo. El formato también es útil para la publicación electrónica y su actualización, y permite informes cuando se hace una consulta informática o mediante una impresión.

Mike Clarke, Murray Enkin, Chris Silagy y Mark Starr desarrollaron el formato original de una revisión Cochrane, con la aportación de muchas otras personas. El formato es suficientemente flexible como para adaptarse a diferentes tipos de revisiones, incluidas las que hacen una única comparación y aquellas que plantean comparaciones múltiples, o a las revisiones preparadas por grupos de ensayos multicéntricos que usan datos de pacientes individuales.

Por múltiples razones, quizá haya quien desee hacer modificaciones del formato de las revisiones Cochrane. Sin embargo, debido al enorme esfuerzo que supondría cambiar la estructura de las revisiones en la Base de datos Cochrane de Revisiones Sistemáticas (*Cochrane Database of Systematic Reviews-CDSR*), el formato debe estar bien definido y ser fijo. Se han hecho algunos cambios menores en el formato descrito en la primera edición del Manual (1994). Estos cambios se han basado en la experiencia de los CRGs, en las aportaciones de los usuarios de las revisiones Cochrane y en sugerencias del Grupo Asesor del programa *Review Manager* (*RevMan*), que ha generado especificaciones para este software que es usado para preparar las revisiones Cochrane. El software *RevMan* está diseñado para ayudar a los revisores a elaborar las revisiones en el formato apropiado y a preparar los archivos requeridos para trasladar las revisiones por medios electrónicos a la Base de Datos Cochrane de Revisiones Sistemáticas (CDSR).

Cada revisión consta de:

- Una carátula donde figuran el título, detalles de citación y direcciones de contacto.
- Una sinopsis – preparada en colaboración con el Centro Cochrane Australiano (*Australasian Cochrane Centre)*
- Un resumen en un formato estructurado.
- El texto de la revisión, que consta de una introducción (antecedentes y objetivos), materiales (criterios de selección y estrategia de búsqueda), métodos, resultados (descripción de los estudios, calidad metodológica y resultados), discusión y conclusiones de los revisores.
- Tablas y figuras estándar para mostrar las características de los estudios incluidos, la especificación de las intervenciones que se compararon, los resultados de los estudios incluidos y una lista de los estudios que se excluyeron.
- Las referencias bibliográficas.

Los encabezados y las tablas estándar ayudan a los revisores cuando preparan su informe y hacen más fácil a los lectores poder identificar la información que sea de su específico interés. Los encabezados se enumeran a continuación. El contenido que debe seguir a cada encabezado se describe en el apéndice 2a (Guía para el Formato de una revisión Cochrane).

2. 1. Esquema de una revisión Cochrane

Carátula:

- Título
- Revisores
- Fuentes de apoyo
- Qué contenido es nuevo

Texto de la revisión

- Sinopsis
- Resumen
  - Antecedentes
  - Objetivos
  - Estrategia de búsqueda
  - Criterios de selección
  - Obtención y análisis de los datos
  - Resultados principales
  - Conclusiones de los revisores
- Antecedentes
- Objetivos
- Criterios para la selección de los estudios de esta revisión
  - Tipos de estudios
  - Tipos de participantes
  - Tipos de intervenciones
  - Tipos de medidas de resultado (*outcomes*)
- Estrategia de búsqueda para la identificación de los estudios
- Métodos de revisión
- Descripción de los estudios
- Calidad metodológica
- Resultados
- Discusión
- Conclusiones de los revisores
  - Implicaciones para la práctica
  - Implicaciones para la investigación
- Agradecimientos
- Conflictos de interés

Referencias:

- Referencias de los estudios
  - Estudios incluidos
  - Estudios excluidos
  - Estudios pendientes de evaluación
  - Estudios en curso
- Otras referencias
  - Referencias adicionales
  - Otras versiones publicadas de esta revisión

Tablas y figuras

- Características de los estudios incluidos
- Características de los estudios excluidos

Características de los estudios en curso
Comparaciones, datos y gráficos
Tablas adicionales

2. 2. Conflictos de Interés.

2. 2. 1. Principio General.

Las revisiones Cochrane deben estar libres de cualquier sesgo real o potencial, causado por la recepción de cualquier beneficio en efectivo o en especies, cualquier patrocinio o subsidio derivados de alguna fuente de financiación que pudiera tener, o que pueda percibirse que tiene un interés en el resultado de la revisión. Es una política de la Colaboración Cochrane que no es aceptable la financiación directa de una fuente individual que tenga intereses particulares en los resultados de la revisión.

2. 2. 2. Recomendaciones

Las siguientes recomendaciones se toman del Código de Conducta de la Colaboración para evitar potenciales conflictos de interés económico (ver apéndice 2b):

- Se debe registrar explícitamente la recepción de beneficios provenientes de cualquier fuente de investigación patrocinada. Deben declararse los conflictos de interés en la CDSR y en otras publicaciones de la Colaboración.
- Si una propuesta plantea un conflicto de intereses grave, debe ponerse en conocimiento del Centro Cochrane local para su valoración (y el Comité Directivo de la Colaboración Cochrane –*Steering Group*- debe ser informado convenientemente).
- No es obligatorio enviar las propuestas de financiación al Centro Cochrane local o al Comité Directivo antes de aceptarlas. Sin embargo, esta supervisión sería aconsejable en los casos de subvenciones selectivas o de cualquier otro tipo que parecieran entrar en conflicto con el principio general expuesto anteriormente.

2. 2. 3. Declaración de los Conflictos de Interés en las revisiones.

Bajo el título "Conflicto de Intereses", los revisores deben informar de cualquier conflicto de intereses capaz de influir en su juicio, incluyendo conflictos personales, políticos, académicos, de tipo financiero o de cualquier índole. Es imposible no tener conflictos de intereses, ya que la única persona que no tiene un interés determinado en un tema, es alguien que no sabe nada acerca de este (Smith 1994). Los conflictos de intereses económicos son los que causan más preocupación; pueden y deben evitarse y debe informarse si hubiese alguno. Cualquier interés secundario (como conflictos personales) que pueda influir indudablemente sobre los juicios que se emiten en una revisión (por ejemplo, en cuanto a la inclusión o exclusión de estudios, la evaluación de la validez de estudios incluidos o la interpretación de los resultados) debe ser reportado.

La declaración de un conflicto de intereses no necesariamente reduce el valor de una revisión y no implica deshonestidad. Sin embargo, los conflictos de intereses pueden influir en los juicios de manera sutil. Los revisores deben poner en conocimiento de los editores del Grupo Colaborador de Revisión (CRG) los potenciales conflictos, aun cuando estén seguros de que sus juicios no han sido o no estarían influenciados. Los editores pueden decidir que no se justifica su declaración, pueden decidir que los lectores deben conocer ese conflicto de intereses para que puedan constituir su propio

criterio acerca de cuán importante es. La decisión acerca de publicar o no tal información debe tomarse conjuntamente por revisores y editores.

Para ayudar a asegurar la integridad real de las revisiones Cochrane y despejar toda duda posible al respecto, todos los revisores deben firmar las siguientes dos afirmaciones además de revelar sus conflictos de interés (ver apéndice 2c):

1) "He participado suficientemente en la concepción y el diseño de esta revisión y en el análisis de los datos, así como en la redacción de la revisión como para asumir la responsabilidad pública por ella. Creo que el manuscrito representa un trabajo válido. He examinado la versión final de la revisión y apruebo su difusión pública".

2) "Certifico que cualquier relación o participación, pasada o presente, en cualquier organización o entidad que tenga un interés económico en el tema o en los materiales tratados en la revisión (por ejemplo, empleo, consultorías, derechos de propiedad, honorarios, testimonio de expertos) se encuentran detalladas a continuación ".

2. 3 Publicación de las revisiones Cochrane en revistas y libros impresos.

Los autores de las revisiones Cochrane pueden desear publicar su trabajo en revistas médicas que tengan una revisión independiente por pares (*peer-review*), en particular en aquellas revistas que hayan expresado su interés en la publicación de revisiones Cochrane.

Para la Colaboración Cochrane, hay una condición de copublicación que es esencial: las revisiones Cochrane deben seguir siendo de difusión gratuita en cualquier medio, sin que nadie pueda imponer restricciones de ningún tipo. Para asegurar esto, los revisores Cochrane conceden una licencia mundial a la Colaboración para estas actividades y no dan derechos de autor exclusivos a ninguna revista o editor. Una revista es libre de solicitar derechos de autor no exclusivos que le permita publicar una revisión una y otra vez, pero esto no puede ser un motivo para restringir la publicación de la revisión en otros formatos que la Colaboración Cochrane considere apropiados.

Se recomienda enfáticamente a los revisores que no publiquen las revisiones Cochrane en revistas antes de que estén listas para ser publicadas en la CDSR. Esto aplica particularmente para los directores de los Centros y para los editores de los Grupos de Revisión. Sin embargo, las revistas a veces pueden insistir en que la publicación de las revisiones en la CDSR no sea anterior a la publicación en papel. Cuando éste es el caso, los revisores deben enviar una revisión para ser publicada en una revista, después de tener el visto bueno del editor de su CRG y antes de su publicación en la CDSR. La publicación en papel no debe estar sujeta a demoras en el tiempo de producción, y los revisores no deben retrasar la publicación de una revisión Cochrane debido a demoras de una revista o con el fin de presentar su revisión en otra revista.

Las revistas también pueden solicitar modificaciones de la revisión por razones editoriales o de contenido. La revisión externa por pares (*peer-review*) a cargo de las revistas puede aumentar el valor de la revisión y debe ser bienvenida.

En general las revistas prefieren revisiones más cortas que aquellas publicadas en la CDSR. Quizá sea apropiado acortar selectivamente las revisiones pero no debe haber ninguna diferencia significativa entre la revisión publicada en las revistas y aquella publicada en la CDSR. Si una revisión se publica en una revista, debe mencionarse que existe una versión más completa que es mantenida en la CDSR. Típicamente

esto debe hacerse incluyendo una declaración como la siguiente en la sección de introducción: “Una revisión más detallada será publicada y actualizada en la Base de Datos Cochrane de Revisiones Sistemáticas (*Cochrane Database of Systematic Reviews).* Referencia”. La referencia debe hacer alusión al protocolo de la revisión publicada en la CDSR. Una declaración similar debe incluirse en la sección de introducción si la revisión es publicada en la CDSR antes de la publicación de la revisión en la revista. Después de que una versión de una revisión Cochrane ha sido publicada en una revista, debe agregarse la referencia de la revista de publicación bajo el encabezado “Otras versiones publicadas de esta revisión”. También se invita a los revisores a que incluyan la siguiente afirmación en las versiones de las revisiones Cochrane que son publicadas en revistas: “Una versión de esta revisión ha sido publicada en la Cochrane Library. Las revisiones sistemáticas Cochrane son regularmente actualizadas para incluir nuevos estudios y en respuesta a los comentarios y las críticas de los lectores. Si usted desea hacer comentarios sobre ésta o sobre otras revisiones Cochrane sobre intervenciones para XXX, por favor envíelos a XXX”. Los Grupos de Revisión pueden querer establecer una política acerca de a cuál persona deben enviarse los comentarios.

Aquellos revisores cuya afiliación primaria es con una entidad Cochrane, deben incluir la siguiente frase al publicar un artículo que no hace referencia a la Colaboración Cochrane o que no refleja la política oficial de esta: “Los puntos de vista expresados en este artículo representan aquellos de los autores y no necesariamente reflejan los puntos de vista o la política oficial de la Colaboración Cochrane”. Adicionalmente, la siguiente modificación en la Cochrane Library debe añadirse a aquellas revisiones Cochrane publicadas en revistas: “Los resultados de una revisión Cochrane pueden interpretarse de manera diferente, dependiendo de las perspectivas de las personas y las situaciones. Por favor, considere cuidadosamente las conclusiones que se presentan. Estas son opiniones de los autores de la revisión y no necesariamente son compartidas por la Colaboración Cochrane”.

El siguiente párrafo puede ser entregado a los editores en el momento de enviar una revisión para su publicación. Debe enviarse copia de la carta de presentación a los editores del GCR para su información. Esta política y procedimiento serán nuevos para la mayoría de los editores de revistas y pueden requerir de una discusión directa con el editor de la revista. Los problemas encontrados en este proceso deben ser notificados a los editores responsables de los CRG.

Se sugiere que se incluya el párrafo siguiente en las cartas de presentación a los editores de la revista:

“*Esta revisión sistemática ha sido preparada bajo el auspicio de la Colaboración Cochrane, una organización internacional que se propone ayudar a tomar decisiones bien fundadas en atención sanitaria al preparar, actualizar y promover el acceso a revisiones sistemáticas sobre los efectos de las intervenciones sanitarias. Las política de publicación de la Colaboración permite a las revistas publicar las revisiones primero si ello fuese procedente, pero también permite a la Colaboración Cochrane publicar y difundir esas revisiones. Las revisiones Cochrane no pueden ser objeto de los derechos de autor exclusivos solicitados por algunas revistas’*”.

2.4. Publicación en formato de revisiones Cochrane de revisiones anteriormente publicadas

La mayoría de las revisiones que han sido previamente publicadas (denominadas como “revisiones anteriormente publicadas”) requieren de un trabajo adicional importante antes de que puedan ser publicadas como una revisión Cochrane en la CDSR. Teniendo en cuenta este trabajo adicional y las sustanciales diferencias en relación con una revisión previamente publicada, la revisión Cochrane puede considerarse como una nueva publicación. La versión previamente publicada de la revisión puede ser referenciada en la revisión Cochrane bajo el encabezado “Otras versiones publicadas de esta revisión”. Sin embargo, en general no es necesario solicitar permiso a la fuente de publicación de la revisión anteriormente publicada.

En ocasiones, una revisión Cochrane puede ser bastante similar a una revisión previamente publicada de manera que el único cambio se refiere al formato de la revisión. En estos casos, los revisores deben obtener permiso con la fuente de publicación de la revisión anteriormente publicada antes de hacerlo en la CDSR. Si los revisores tienen dudas acerca de si deben solicitar este permiso, se invita a que de todas maneras lo hagan. Es poco probable que esto plantee problemas, contando con que esto se haga antes de planear su presentación en la CDSR. Si se sabe de antemano que hay un interés en publicar una versión de una revisión Cochrane en una revista, los revisores no deben asignar ninguna cláusula de derechos de autor a la revista (ver sección 2.3). La Colaboración Cochrane no requiere la cesión de derechos exclusivos de autor. Por lo tanto, no hay ningún problema en publicar una versión de una revisión Cochrane en una revista después de que ésta ha sido publicada en la CDSR, siempre y cuando no sea denominada como revisión Cochrane y se reconozca que se basa en una revisión Cochrane (ver sección 2.3.).

La transformación de revisiones con datos individuales de pacientes en revisiones Cochrane se discute en la sección 11.6.

Referencias

Smith 1994. Smith R. Conflict of interest and the BMJ. BMJ 1994; 308: 4-5.

3. DESARROLLO DE UN PROTOCOLO.

3. 1. Justificación de los protocolos.

La preparación de una revisión es un proceso complejo que implica muchas consideraciones y decisiones acerca del proceso y los recursos que son necesarios (ver apéndice 3a). Como en cualquier proyecto científico, los métodos que se van utilizar deben establecerse de antemano. Sin embargo, las revisiones son por naturaleza de carácter retrospectivo dado que los estudios incluidos habitualmente se identifican una vez ya han sido completados y publicados. Por lo tanto, es importante que el proceso a seguir sea tan riguroso y haya sido tan bien definido como sea posible (Light 1984b), manteniendo a su vez una perspectiva práctica. El conocimiento de los resultados del estudio por parte del revisor puede tener una influencia sobre:

- La definición de la pregunta de una revisión sistemática.
- Los criterios para la selección de los estudios.
- Las comparaciones para los análisis.
- Los resultados que se van a reportar en la revisión.

Así como, a veces, los protocolos de los ensayos clínicos aleatorizados deben ser modificados para adaptarse a circunstancias imprevistas (como, por ejemplo, problemas en el reclutamiento de los participantes, en la obtención de los datos o por unas tasas inesperadas de los eventos estudiados), a veces también resulta necesario introducir cambios en un protocolo de revisión. Aunque es necesario hacer todos los esfuerzos posibles para adherirse al protocolo que ha sido establecido de antemano, debe reconocerse que esto no siempre resulta posible o apropiado. Los cambios en el protocolo no deben hacerse en función de la manera como éstos afectan los resultados de la revisión. Las decisiones *a posteriori* (como sería la exclusión de estudios previamente seleccionados) que se toman a sabiendas de su repercusión sobre los resultados de la revisión deben evitarse ya que, con mucha probabilidad, pueden introducir sesgos. Por norma, todos los cambios introducidos en los protocolos deben estar bien documentados y reportados y, siempre que sea posible, deben llevarse a cabo "análisis de sensibilidad" (ver sección 8. 10) para analizar el impacto que tales decisiones tienen sobre los resultados de la revisión.

El protocolo de una revisión Cochrane debe consistir de las siguientes secciones:

- Carátula.
- Antecedentes.
- Objetivos.
- Criterios de selección de los estudios de esta revisión.
- Estrategia de búsqueda para la identificación de los estudios.
- Agradecimientos.
- Conflictos de interés.
- Métodos de la revisión.
- Referencias adicionales.
- Tablas adicionales.

En el apéndice de la sección 2 (Guía sobre el formato de una revisión Cochrane) se describe el contenido que deben tener cada una de estas secciones en una revisión completa. En las secciones 4, 5, 6, 7 y 8 se explicitan los criterios metodológicos a tener

en cuenta para decisiones específicas referentes a la formulación del problema; la identificación, selección y evaluación de los estudios; la obtención y análisis de los datos.

3. 2. Los antecedentes de una revisión.

Las preguntas bien formuladas habitualmente no surgen de la nada, sino que se producen en el contexto de un cuerpo de conocimiento ya formado. Este contexto debe ser descrito en la sección de los antecedentes de la revisión. Puede incluir información acerca de aspectos relacionados con la biología, la epidemiología, la importancia desde el punto de vista de la salud pública, la relevancia clínica o la práctica actual relacionada con el tema que serán estudiados en la revisión. Debe hacerse referencia a cualquier otra revisión sistemática que haya sido previamente publicada y que trate la misma cuestión, o bien a la existencia de controversias en torno al tema. Estos antecedentes ayudan a establecer la justificación de la revisión, y deben explicar por qué las preguntas que se plantean son importantes. Los antecedentes y los objetivos, junto con la estrategia de búsqueda propuesta y los planes para recolectar y analizar los datos, constituyen la base del protocolo de una revisión Cochrane. Los editores de los Grupos Colaboradores de Revisión (Collaborative Review Groups) evalúan y proporcionan asesoramiento sobre estos protocolos antes de que las revisiones propiamente dichas se lleven a cabo. El protocolo también será publicado en la Base de Datos Cochrane de Revisiones Sistemáticas (Cochrane Database of Systematic Reviews*)* y pueden ser objeto de comentarios y críticas por parte de sus usuarios.

En general, las revisiones sistemáticas pueden verse motivadas por 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. Aunque las revisiones Cochrane puedan estar motivadas por cualquiera de estos u otros factores, su propósito principal debería ser el de resumir la evidencia científica y ayudar a las personas a entenderla. Los revisores deben tener cuidado en no imponer sus propios valores y preferencias sobre los de los demás, cuando responden las preguntas que se plantean. Deben ayudar a la gente a tomar decisiones sanitarias que sean prácticas. Esto tiene implicaciones importantes a la hora de decidir si debemos emprender o no una revisión Cochrane, cómo formular el problema que ésta abordará, cómo desarrollar el protocolo y cómo presentar los resultados de la misma.

- Las preguntas deben abordar las diferentes alternativas (opciones prácticas) a las que nos enfrentamos cuando tomamos decisiones sanitarias.
- Las revisiones deben tener en cuenta las variables de resultado (*outcomes*) que sean significativas o relevantes para aquellas personas que deben tomar las decisiones.
- Los revisores deben describir cómo abordarán tanto los efectos adversos como los beneficios.
- Los métodos a utilizar en una revisión deben ser aquellos que permitan optimizar las probabilidades de que los resultados proporcionen la mejor evidencia disponible sobre la que basar las decisiones.
- Es importante que los lectores puedan conocer cuándo no existen evidencias fiables, o si éstas no existen con relación a variables de resultado específicas que probablemente sean de importancia para aquellos que toman las decisiones.
- No resulta de utilidad incluir evidencia para la cual existe un alto riesgo de estar afectada por sesgos dentro de una revisión, aún cuando no exista otra mejor evidencia disponible.

- 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 de los estudios originales seleccionaron para evaluar los efectos de la intervención.
- Siempre que sea posible, es importante mantener una perspectiva internacional. Las evidencias científicas a considerar por la revisión no deben quedar restringidas por un criterio de nacionalidad o de idioma sin una razón de peso que lo justifique.

Cuando el protocolo se convierte en una revisión completa, debe mencionarse que la revisión estaba precedida de la publicación de un protocolo. Es una política de la Colaboración que los protocolos que no han sido convertidos en revisiones completas al cabo de dos años deben, en general, ser excluidos de la CDSR.

3.3. Referencias

Light 1984b. Light RJ, Pillemer DB. Organizing a reviewing strategy. In: Summing Up: The Science of Reviewing Research. Cambridge, Massachusetts: Harvard University Press, 1984;13-31

4. FORMULACIÓN DEL PROBLEMA.

4. 1. Justificación de las preguntas bien formuladas.

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.

Al igual que en cualquier proyecto de investigación, la primera y más importante decisión a la hora de preparar una revisión es determinar el enfoque de la misma (Light 1984b). Esto se consigue mejor planteando preguntas claramente delimitadas. Tales preguntas resultan esenciales a la hora de determinar la estructura de una revisión (Jackson 1980, Cooper 1984, Hedges 1994). Concretamente, guiarán gran parte del proceso de revisión como son 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. Los lectores utilizan las preguntas y los objetivos planteados para juzgar si la revisión les puede resultar de interés y si ésta es pertinente para las cuestiones que ellos se plantean.

4. 2. Componentes clave de una pregunta.

Existen varios componentes clave que son necesarios para formular bien una pregunta (Richardson 1995, Counsell 1997), y que deben especificarse en la sección de “Criterios de selección de los estudios” de la revisión. Una pregunta claramente definida debe especificar el tipo de personas (participantes), el tipo de intervenciones o de exposiciones y el tipo de 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. En general, cuanto más precisa sea la definición de los componentes, más focalizada resultará la revisión. No es necesario el mismo grado de precisión para definir cada uno de los componentes clave. Por ejemplo, uno quizá desee centrarse en los diversos tratamientos disponibles para un estadio específico del cáncer de mama, o por el contrario, centrarse en una droga en particular para cualquier estadio del cáncer de mama. En el primer ejemplo, en la sección “Tipos de participantes” tendría que definirse con toda precisión el estadio y la gravedad de la enfermedad, mientras que en el segundo ejemplo, debería definirse con extrema precisión el tratamiento propuesto dentro de la sección “Tipos de Intervención”.

A continuación, se presentan los componentes clave de la pregunta, con ejemplos de aspectos útiles que cabe considerar en cada uno de ellos. Los revisores deben estar familiarizados con los términos y el lenguaje empleados en distintos lugares y contextos para describir estos mismos componentes.

4. 2. 1. ¿Qué tipos de personas (participantes)?

A menudo, resulta de utilidad considerar en dos fases el tipo de personas que son objeto de interés para la revisión. En primer lugar, hay que definir las enfermedades o condiciones que son de interés. Deben establecerse criterios que sean suficientemente explícitos para determinar la presencia de la enfermedad o de la condición de interés. En segundo lugar, hay que identificar la población y el ámbito de interés. 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, raza, nivel de instrucción

o la presencia de una condición particular como la angina o disnea. Quizá uno también pueda estar interesado en un entorno o ámbito específico en función de si las personas están viviendo en la comunidad, están hospitalizadas, en centros de convalecencia o instituciones de atención para crónicos, o bien si se trata de pacientes en régimen ambulatorio.

Cualquier restricción basada en características específicas de la población o del ámbito de estudio debe estar justificada por evidencia ya reconocida. Por ejemplo, enfocar una revisión sobre la efectividad del *screening* (cribaje, tamizaje) de cáncer de mama mediante mamografía selectiva destinada a mujeres entre los 40 y 50 años de edad puede estar justificado por razones de tipo biológico, revisiones sistemáticas publicadas con anterioridad o por la controversia existente. Por el contrario, debe evitarse centrar una revisión seleccionando un sub-grupo particular de personas, en función de su edad, sexo o signo zodiacal al nacer, simplemente por intereses personales, cuando no exista ninguna justificación subyacente de tipo biológico o sociológico para ello. Cuando hay incertidumbre acerca de la posible existencia de diferencias importantes del efecto observado entre diversos sub-grupos de la población, probablemente sea mejor incluir todos los sub-grupos relevantes dentro del análisis y, luego, examinar en qué medida esas diferencias sobre el efecto son importantes y plausibles en el análisis (ver secciones 4.5 y 8 más adelante).

4. 2. 2. ¿Qué tipos de comparaciones (intervenciones)?

El siguiente componente clave de una pregunta bien formulada es especificar las intervenciones que son de interés. También es importante definir las intervenciones contra las que éstas serán comparadas, es decir, el tipo de grupos control que son aceptables para la revisión. Debe considerarse con detenimiento si las personas en el grupo control pueden estar recibiendo otras intervenciones distintas al placebo y si aquellas intervenciones se superponen de alguna manera con la intervención activa que está siendo evaluada. Esta cuestión se discute en detalle en la sección correspondiente a la evaluación de la calidad de los estudios (sección 6).

4. 2. 3. ¿Qué tipos de resultados?

El tercer componente clave de una pregunta bien formulada es la especificación de aquellos resultados (*outcomes*) concretos que sean de interés. Aunque en las revisiones Cochrane deben incluirse todos los resultados que sean importantes, no deben incluirse aquellos que sean considerados triviales. Los revisores deben evitar abrumar a los lectores con datos que tienen poca o ninguna importancia. Al mismo tiempo, deben tener cuidado en no omitir datos importantes. Siempre que sea necesario, hay que definir los criterios explícitos para establecer la presencia de dichos resultados. De igual manera, deberá especificarse si se tendrán en cuenta resultados que se presentan combinados. Por ejemplo, si un estudio solamente proporciona datos combinados de accidentes cerebrovasculares mortales y no mortales, ¿será incluido si la pregunta se refiere específicamente a las muertes por accidente cerebrovascular?

En general, las revisiones Cochrane deben incluir todos aquellos resultados publicados que puedan ser significativos para las personas que necesitan tomar alguna decisión acerca del problema sanitario del que se ocupa la revisión. Aún más, resulta importante especificar aquellos resultados que son más trascendentes para los responsables de tomar decisiones, aún cuando sea improbable que se vayan a encontrar esos datos. Por ejemplo, la calidad de vida es un resultado importante -quizá el más importante- para las personas que deben decidir sobre el empleo o no de quimioterapia para el cáncer avanzado, a pesar de que los estudios disponibles habitualmente sólo reportan datos de supervivencia. Además, los revisores deben indicar cómo intentarán incluir datos sobre

efectos adversos en su revisión. Con respecto a este punto, en lugar de incluir una lista exhaustiva de todos los efectos adversos, quizá sea más informativo resumirlos en resultados adversos "graves" (lo suficientemente graves como para obligar al abandono del tratamiento) y menores, incluyendo una descripción adecuada de ellos.

A veces es posible conseguir datos no publicados de los mismos investigadores con el fin de separar aquellos resultados que se presentan combinados, y también con otras finalidades (ver sección 7). Antes de decidir excluir un estudio que parece cumplir los criterios de selección pero que no reporta los resultados de una manera que sea adecuada para la revisión, vale la pena considerar la posibilidad de tratar de obtener la información necesaria contactando directamente con sus investigadores.

4. 2. 4. ¿Qué tipos de diseños de estudio?

Algunos diseños de estudio resultan superiores a otros para tratar de responder a determinadas preguntas. Muchos consideran que los ensayos clínicos controlados y aleatorizados son una 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 pueden ser abordadas mediante estudios de cohortes o de casos y controles. Los revisores deberán considerar en cada caso qué tipos de diseños de estudio pueden proporcionarnos datos fiables con los que responder a las preguntas planteadas.

Otros aspectos relevantes acerca del diseño de los estudios que merecen ser considerados desde el inicio son si se revisarán aquellos estudios que: tienen un grupo control con placebo, que evalúan los resultados de una manera imparcial o libre de sesgos, o que tienen un seguimiento de cierta duración. Cuanto más restrictivos sean los revisores en los criterios exigidos acerca del diseño de los estudios para responder a la pregunta planteada, menos probabilidades tendrán de encontrar datos que respondan de forma específica a preguntas tan restringidas.

No obstante, la revisión de estudios que proporcionan datos poco fiables con los que responder a una pregunta, no sólo representa una pérdida de tiempo sino que puede conducirnos a conclusiones erróneas. Por ejemplo, si uno está interesado en saber si una terapia mejora la supervivencia en pacientes con una enfermedad crónica, podría ser inapropiado analizar estudios de una duración muy corta, excepto para dejar bien patente que éstos no resultan adecuados para abordar la cuestión de interés planteada.

Ya que las revisiones Cochrane plantean cuestiones acerca de los efectos de la atención sanitaria, éstas se centran principalmente en ensayos clínicos aleatorizados. Existen dos razones por las que uno debe ser cauteloso a la hora de incluir estudios no aleatorizados en una revisión sobre los efectos de la atención sanitaria, ambos relacionados con los sesgos. En primer lugar, aunque es posible controlar los factores de confusión que son conocidos y mesurables mediante otro tipo de diseños de estudio, la aleatorización es la única manera de controlar aquéllos que no se conocen o no están cuantificados. En las intervenciones clínicas, la decisión acerca de quién debe recibir una intervención y quién no está influida por muchos factores, incluidos los factores pronósticos. Datos empíricos indican que, por lo general, los estudios no aleatorizados tienden a sobrestimar los efectos de las intervenciones sanitarias (Sacks 1982, Chalmers 1983, Schulz 1995). Sin embargo, una revisión sistemática sobre métodos ha demostrado que el grado e incluso la dirección del sesgo en los estudios no aleatorizados resultan a menudo imposibles de predecir (Kunz 1998).

En segundo lugar, aunque a menudo es difícil localizar ensayos clínicos aleatorizados (Dickersin 1994) y las revisiones que dejan de incluir ensayos clínicos no publicados pueden estar sesgadas al sobrestimar la efectividad de una intervención (Dickersin 1993), los esfuerzos de la Colaboración Cochrane dirigidos a identificar ensayos clínicos aleatorizados no se han visto acompañados por esfuerzos similares para identificar otros tipos de estudio. En consecuencia, la inclusión dentro de una revisión de otros estudios que no sean ensayos clínicos controlados y aleatorizados puede requerir de esfuerzos adicionales para identificarlos y para actualizar la revisión, lo que puede aumentar el riesgo de que los resultados de la revisión estén afectados por un sesgo de publicación.

A pesar de las consideraciones anteriores, a veces resulta apropiado realizar una revisión sistemática de los estudios no aleatorizados sobre los efectos de la atención sanitaria. Por ejemplo, en ocasiones el curso de una enfermedad es tan uniforme o los efectos de una intervención son tan evidentes que resulta innecesario e incluso no ético realizar un ensayo clínico aleatorizado. En tales circunstancias sería insensato restringir una revisión sólo a ensayos clínicos aleatorizados. Aunque siempre resulta fundamental tratar de evitar cualquier tipo de sesgos decidiendo qué tipos de estudios pueden incluirse en la revisión, son los revisores individuales y los Grupos Colaboradores de Revisión quienes deben decidir qué tipos de estudios son los más convenientes para abordar las cuestiones específicas planteadas.

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

Una vez que uno tiene una pregunta bien planteada, debe determinar qué componentes clave tendrá en cuenta para definir la estrategia de búsqueda inicial. Para las revisiones Cochrane, la búsqueda de estudios se ve enormemente facilitada por la existencia de los registros especializados de ensayos clínicos complicados por los Grupos Colaboradores de Revisión. Sin embargo, el grado de desarrollo de los registros varía y quizá esto haga necesario que los revisores deban llevar a cabo búsquedas suplementarias.

Las búsquedas que requieran de la presencia simultánea de varios de los componentes de la pregunta -o bien de formulaciones muy específicas de ciertos componentes- suelen ser demasiado específicas y tienden a pasar por alto información relevante. Por ejemplo, si uno busca estudios que aborden los efectos a largo plazo de la insulinoterapia sobre la función renal en diabéticos tipo II y lo hace buscando aquéllos que están indexados mediante los términos "diabetes tipo II", "insulina", "función renal" y "largo plazo", es probable que no se identifiquen estudios que pueden ser relevantes. Por otro lado, si sólo se emplea los términos "insulina" o "diabetes tipo II" como términos de búsqueda, es probable que se identifiquen cientos de artículos irrelevantes.

En general, los componentes clave que resultan de utilidad cuando se hace una búsqueda incluyen la condición o enfermedad de interés y la intervención o exposición que se está evaluando. Aunque uno quizá esté específicamente interesado en un ámbito determinado, los estudios no suelen estar indexados según el tipo de ámbito dentro de las bases de datos electrónicas. Además, los estudios pueden evaluar múltiples resultados, algunos de los cuales pueden ser relevantes para la revisión, pero pueden no formar parte en la indexación del artículo. Este asunto se discute con detalle en la siguiente sección sobre localización y selección de estudios (sección 5).

Cualquiera que sean las estrategias de búsqueda utilizadas, será necesario tener que revisar numerosos 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 nos obligará a especificar los criterios que se utilizarán para seleccionar los estudios. Sin embargo, a menudo se requiere de una mayor clarificación de los criterios de selección para desarrollar reglas de decisión que sean sensibles y reproducibles. Si, por ejemplo, uno está revisando estudios sobre el tratamiento del estreñimiento, deberá decidir si revisará los estudios que abordan el estreñimiento agudo y/o crónico, así como establecer criterios aceptables de definición de estreñimiento agudo y crónico. ¿Está usted interesado en todo tipo o grado de estreñimiento, o sólo en el estreñimiento grave? En tal caso, ¿cómo se define el estreñimiento "grave"? ¿Revisará los estudios que definen el estreñimiento en función de la frecuencia deposicional semanal o bien se limitará a los estudios que definen el estreñimiento en función de la presencia de síntomas tales como el esfuerzo para evacuar o la consistencia de las heces? ¿Examinará sólo los estudios que hayan establecido el mecanismo fisiopatológico subyacente del estreñimiento o se limitará a revisar ciertos trastornos fisiopatológicos específicos? ¿Tendrá en cuenta usted los estudios que simplemente afirman que los participantes estaban "estreñidos"?

4. 4. Utilización de los componentes clave de la pregunta para guiar la recopilación de los datos.

De cada uno de los estudios individuales, los revisores registrarán aquellos datos relevantes en relación con los componentes clave de la pregunta. Por tanto, las preguntas bien formuladas están relacionadas directamente con el proceso de recopilación de datos porque guían: el establecimiento de los criterios finales que se utilizarán para seleccionar los estudios que resultan adecuados para la revisión; y los datos que deben ser extraídos de los estudios que satisfacen aquellos criterios de selección.

Los componentes clave de la pregunta también pueden estar directamente relacionados con la manera en la que se van a presentar y analizar los datos. Estos temas serán discutidos en las secciones 6, 7 y 8.

4. 5. ¿Preguntas amplias o preguntas restringidas?

El alcance de las preguntas que aborda una revisión puede ser amplio o restringido. Por ejemplo, un revisor podría abordar una pregunta amplia al considerar si los agentes antiplaquetarios en general son eficaces en la prevención de los eventos trombóticos en seres humanos. Por el contrario, un revisor podría preguntarse si un agente antiplaquetario específico – como la aspirina- es efectivo para reducir el riesgo de sufrir un proceso trombótico específico -el accidente cerebrovascular- en personas ancianas con antecedentes de este tipo de patología.

Determinar el alcance de un tema objeto de revisión es una decisión que va a depender de múltiples factores, incluida la visión que se tenga sobre su relevancia y su potencial impacto; la información teórica, biológica y epidemiológica que la sostiene; la posibilidad potencial de generalización y de obtener respuestas válidas y de los recursos disponibles.

Son varias las ventajas y desventajas que se derivan del hecho de formular preguntas amplias o restringidas. Las revisiones enfocadas de manera restringida quizá no sean generalizables a entornos diversos, a distintas poblaciones o a formas alternativas de una misma intervención. También pueden dar lugar a conclusiones espurias o sesgadas del mismo modo que, a veces, ocurre con el análisis de sub-grupos (ver sección 8.7). Por ejemplo, una revisión sobre la eficacia de la aspirina para prevenir accidentes cerebrovasculares en mujeres podría conducir a la falsa conclusión de que la aspirina no fue eficaz en las mujeres, cuando en realidad no había suficientes datos para detectar

ninguna diferencia de efecto significativa entre hombres y mujeres. Un enfoque restringido conlleva un elevado riesgo de conclusiones sesgadas cuando el revisor conoce la bibliografía sobre un determinado tema y restringe los criterios de inclusión de tal manera que queden excluidos algunos o varios estudios con resultados que estén en conflicto con la postura que defiende el revisor. También existe el peligro de que un conocimiento previo de los resultados derivados de una serie de estudios sobre un tipo de intervenciones pueda influir sobre la elección de una intervención específica de este mismo tipo para plantear una revisión restringida.

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

La búsqueda de datos que sean relevantes para las preguntas muy amplias resulta más lenta y costosa que la búsqueda sobre temas restringidos. Ya que las preguntas amplias suelen abordarse mediante grandes cantidades de estudios muy heterogéneos, la síntesis e interpretación de los datos puede convertirse en un auténtico desafío. Las revisiones con un enfoque amplio también pueden tornarse inmanejables para su presentación, mantenimiento y comprensión.

Una opción que se considera útil consiste en plantear una revisión de enfoque amplio a partir de una serie de revisiones más restringidas. Por ejemplo, los proveedores de servicios sanitarios y las mujeres embarazadas que quieren dejar de fumar son más proclives a querer saber qué estrategia debería utilizarse para abandonar el hábito de fumar (una pregunta amplia). Una revisión que les ayudaría a responder a esta pregunta podría plantearse a partir de una serie de revisiones más reducidas que se centren cada una de ellas en la eficacia que tiene una estrategia específica, como podría ser la modificación del comportamiento.

Tanto si se comienza con preguntas más restringidas y luego se amplía planteando una pregunta de rango superior, como si se empieza por una cuestión amplia y se divide en cuestiones menores, todo depende de la naturaleza del problema (por ejemplo, cuán complejo sea éste, si está bien comprendido, cuánta investigación se ha realizado sobre el tema) y de las circunstancias particulares de los revisores o de su Grupo Colaborador de Revisión (como por ejemplo, el grado de desarrollo de su registro especializado de ensayos clínicos, la posibilidad de obtener recursos, el tiempo e interés).

4. 6. Modificación de la pregunta.

Aunque las preguntas deben plantearse en el protocolo antes de iniciar la revisión completa, estas preguntas no deben convertirse en una camisa de fuerza que impida la exploración de aspectos inesperados (NHS CRD 1996). Las revisiones son análisis de datos ya existentes que se refieren a poblaciones, entornos, intervenciones, medidas de resultado y diseños de estudio que han sido elegidos de antemano. Generalmente no es posible formular una pregunta para una revisión sin conocer algunos de los estudios relevantes sobre el tema planteado, y puede darse el caso de que se plantee la necesidad de modificar la pregunta que aborda la revisión a la luz de la evidencia acumulada en el proceso de realización de la misma.

Aunque es de esperar que se obtenga una cierta fluidez y refinamiento al formular las preguntas de revisión conforme se va adquiriendo una comprensión más plena del problema, es importante protegerse del riesgo de sesgo al modificar las preguntas. Las preguntas *a posteriori* son más proclives al sesgo que aquellas que han sido formuladas previamente, y las preguntas modificadas por los propios datos pueden generar conclusiones falsas basadas en resultados espurios. Debe documentarse cualquier cambio que sufra el protocolo derivado de la modificación de la pregunta inicial de revisión. Cuando se modifican o refinan las preguntas es útil hacerse las siguientes preguntas:

- ¿Cuál es el motivo del cambio?
- ¿Se hizo después de que uno viera y fuera influido por los resultados de un estudio en particular? ¿O sencillamente fue porque no se habían considerado inicialmente otras formas alternativas -aunque aceptables- de definir a los participantes, las intervenciones y/o los resultados de interés?
- ¿Resultan las estrategias de búsqueda apropiadas para la pregunta que ha sido refinada (especialmente cuando ya han sido ejecutadas)?
- ¿Se adapta el proceso de extracción de datos a esta pregunta que ha sido refinada?

4.7. Referencias

Chalmers 1983. Chalmers TC, Celano P, Sacks HS, Smith H, Jr. Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983; 309:1358-61.

Cooper 1984. Cooper HM. The problem formulation stage. In: Cooper HM, editor. Integrating Research. A Guide for Literature Reviews. Newbury Park: Sage Publications, 1984; 19-37.

Counsell 1997. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. Ann Intern Med 1997; 127: 380-7.

Dickersin 1993. Dickersin K, Min YI. NIH clinical trials and publication bias. Online J Curr Clin Trials [serial online] 1993; Doc No 50.

Dickersin 1994. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. BMJ 1994; 309:1286-91.

Hedges 1994. Hedges LV. Statistical considerations. In: Cooper H, Hedges LV, editors. The Handbook of Research Synthesis. New York: Russell Sage Foundation, 1994; 30-3.

Jackson 1980. Jackson GB. Methods for integrative reviews. Rev Educ Res 1980; 50:438-60.

Kunz 1998. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised trials. BMJ 1998; 317: 1185-90.

Light 1984. Light RJ, Pillemer DB. Organizing a reviewing strategy. In: Summing Up: The Science of Reviewing Research. Cambridge, Massachusetts: Harvard University Press, 1984; 13-31.

NHS CRD 1996. NHS Centre for Reviews and Dissemination. Undertaking Systematic Reviews of Research on Effectiveness (CRD Report 4). York: The University of York, 1996.

Richardson 1995. Richardson WS, Wilson MS, Nishikawa J, Hayward RSA. The Well-built Clinical Question: A Key to Evidence Based Decisions. ACP J Club 1995; A12-3.

Sacks 1982. Sacks HS, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. Am J Med 1982; 72:233-40.

Schulz 1995. 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.

5. LOCALIZACIÓN Y SELECCIÓN DE LOS ESTUDIOS.

Que la búsqueda de los estudios sea exhaustiva 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 identificar entre un 30% y un 80% de todos los ensayos controlados aleatorios que han sido publicados (dependiendo del área o del tema específicos). Incluso aunque los registros relevantes estén en MEDLINE puede no ser fácil su identificación. Por otro lado hay que tener en cuenta que esta base de datos no incluye la totalidad de la literatura médica publicada, y que, además, algunos estudios nunca llegan a publicarse (Dickersin 1994). Las referencias de estudios escritos en lengua no inglesa tienen una baja representación en MEDLINE, y sólo están incluidos los artículos publicados, por lo que si se utiliza esta base de datos para identificar estudios, una revisión puede estar afectada potencialmente por un sesgo de publicación (es decir que se seleccionan prioritariamente para ser publicados estudios con resultados positivos) (Simes 1986, Dickersin 1987, Simes 1987, Begg 1988, Hetherington 1989, Easterbrook 1991, Dickersin 1992b, Dickersin 1993, Dickersin 1994) y por un sesgo idiomático (Grégoire 1995, Moher 1996a, Egger 1997). Para evitar estos sesgos y asegurar que, en la medida de lo posible, todos los datos relevantes sean incluidos en una revisión es necesario utilizar varias fuentes de información para identificar los estudios y realizar sistemáticamente la selección de aquellos que se van a incluir en la revisión. Debe tenerse en cuenta la posibilidad de que haya un sesgo de referencias (una tendencia a citar preferentemente aquellos estudios que apoyan el propio punto de vista) y evitarlo mediante el empleo de múltiples estrategias de búsqueda (Gotzsche 1987, Ravnskov 1992).

En esta sección trataremos en primer lugar las fuentes de información necesarias para identificar los estudios, incluidos los registros especializados de cada Grupo Colaborador de Revisión (CRG), la verificación de listas de referencias, las comunicaciones personales, las bases de datos electrónicas y la búsqueda manual. Seguidamente se abordará el desarrollo de una estrategia de búsqueda, la selección de los estudios, cómo documentar una estrategia de búsqueda, cómo hacer un seguimiento de los estudios y finalmente, cómo contribuir al registro de ensayos.

5. 1. Registros especializados de los Grupos Colaboradores de Revisión - CRG.

De acuerdo con los principios que guían el desarrollo de la Colaboración Cochrane, cada CRG debe procurar satisfacer su necesidad específica de identificación de estudios - dentro del ámbito de sus objetivos - de la manera más exhaustiva posible, protegiéndose contra los sesgos y reduciendo al mínimo la duplicación de esfuerzos. En general, esto se hace mediante el desarrollo y el mantenimiento de los registros de estudios. Depende de cada CRG la decisión de cuales son los mecanismos más eficientes y su implementación para asegurar que los estudios de su registro especializado sean accesibles al realizar revisiones. El equipo editorial asumirá alguna responsabilidad, si no toda, en la identificación de estudios nuevos y los remitirá a los correspondientes revisores, quienes deberán evaluar si reúnen los criterios de inclusión para sus revisiones. En algunos casos, se puede pedir a los propios revisores que comprueben cómo funciona el registro especializado del CRG, o pueden utilizarse otros abordajes alternativos. En la revisión debe informarse del mecanismo utilizado, incluyendo la última fecha en que se comprobó el funcionamiento del registro especializado del CRG y si se identificaron estudios potencialmente elegibles. Como mínimo una vez al año, dependiendo de la naturaleza del tema,

debería hacerse una búsqueda para cada revisión en bases de datos como CENTRAL/CCTR, con el objetivo de identificar nuevos estudios, (ver más adelante).

La principal fuente de estudios para las revisiones Cochrane es el registro especializado del CRG. Hay diferencias importantes sobre cómo (y hasta qué nivel) los CRG han desarrollado sus registros (ver detalle de los CRG en la Cochrane Library. Independientemente, los revisores comparten la responsabilidad, junto con su equipo editorial, de asegurar la inclusión en las revisiones de tantos estudios válidos como sea posible identificar y también deben considerar la posibilidad de usar estrategias adicionales para asegurar que todos los estudios pertinentes queden registrados. Al desarrollar una estrategia de búsqueda es importante consultar al coordinador de Búsqueda de Ensayos o al coordinador del grupo de revisión del CRG para evitar la duplicación innecesaria de esfuerzos y garantizar que se está haciendo un uso eficiente de otras bases de datos adicionales. Los CRG deben esforzarse por obtener copias de los artículos de su registro de modo que los revisores que los consideren relevantes puedan acceder a ellos.

5. 2. Control de las listas de referencia.

Los revisores deben controlar las listas de referencias de todos los artículos relevantes que hayan identificado (incluyendo las referencias de revisiones sistemáticas publicadas con anterioridad). En general, éste es un medio eficiente, aunque lento, de identificar estudios. El proceso de rastrear las referencias de un artículo a otro se denomina "buscar perlas" (*pearling*) método de la ascendencia o de concatenación de citas. Además, deben obtenerse los artículos potencialmente relevantes que sean identificados para evaluar su posible inclusión en la revisión.

Algunas de las fuentes de referencia de ensayos más obvias y convenientes son las revisiones ya existentes. Deben obtenerse y examinarse las copias de las revisiones sobre el tema de interés publicadas con anterioridad para obtener las referencias de los estudios originales. La Cochrane Library incluye la Base de Datos de Resúmenes de Revisiones de Efectividad (*Database of Abstracts of Reviews of Effectiveness-DARE*), producida por el *NHS Centre for Reviews and Dissemination (CRD)* en York, RU.. La DARE proporciona información acerca de las revisiones publicadas anteriormente sobre los efectos de la atención sanitaria, que pueden ser utilizadas para identificar estudios relevantes. La DARE incluye dos tipos principales de registros:

- Resúmenes estructurados que evalúan y resumen las revisiones sistemáticas publicadas con anterioridad y consideradas de buena calidad.
- Registros de referencias: referencias a otras revisiones sistemáticas publicadas. En general, la calidad de estos registros no ha sido evaluada por el CRD.

MEDLINE, EMBASE y otras bases de datos bibliográficas también pueden usarse para identificar artículos de revisión. En MEDLINE, los artículos de revisión aparecen indexados como *Review* en el campo *Publication Type* Se han desarrollado estrategias de búsqueda para mejorar la identificación de este tipo de publicaciones (Boynton 1998).

5. 3. Comunicación personal.

Los colegas pueden ser una importante fuente de información de ensayos recientes que todavía no hayan sido publicados o de ensayos inéditos más antiguos. A veces estos canales informales de comunicación pueden ser los únicos medios para identificar datos no publicados. La "literatura gris", compuesta por informes técnicos,

actas de congresos y tesis doctorales, entre otros, también puede resultar útil, ya que muchos estudios referenciados de esta manera no son publicados en su totalidad (Scherer 1994). También pueden usarse cartas formales de solicitud de información con el objeto de identificar estudios. Una manera de hacerlo es enviar una lista exhaustiva de artículos relevantes para la revisión junto con los criterios de inclusión dirigida al primer autor de los artículos incluidos y preguntarle si conoce algún estudio adicional que pueda ser relevante, publicado o no. También puede ser aconsejable enviar la misma carta a otros expertos, empresas farmacéuticas o a otros sujetos que tengan interés en el tema.

5. 4. Bases de datos electrónicas.

5. 4. 1. Registro Cochrane de Ensayos Controlados

Desde Octubre de 1966, la Colaboración Cochrane ha venido desarrollando su propio registro central de estudios que posiblemente sean relevantes para las revisiones Cochrane; el registro se denomina "CENTRAL". Este registro pretende incluir todas las citas relevantes que han sido identificadas mediante el trabajo desarrollado por la Colaboración Cochrane y posteriormente transferidas y procesadas por la *Providence Office* (NECC@P) del *New England Cochrane Center,* antes llamado *Baltimore Cochrane Center.* CENTRAL se publica en la Cochrane Library dentro del *Cochrane Controlled Trials Register* (CENTRAL/CCTR). Incluye principalmente referencias de los ensayos controlados aleatorios y ensayos clínicos controlados en atención sanitaria. 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 indexación de CENTRAL/CCTR evolucionan para mejorar su utilidad como fuente principal para la identificación de ensayos. La meta a largo plazo es que CENTRAL sirva a los CRG como una fuente permanentemente actualizada de ensayos controlados y de los contenidos de registros especializados i que CCTR sea un producto más refinado, de manera que incluya únicamente ensayos controlados. En la Cochrane Library hay información adicional al respecto.

5. 4. 2. MEDLINE y EMBASE.

Index Medicus (publicado por la *US National Library of Medicine (NLM))* y Excerpta Medica (publicada por Elsevier) son repertorios de revistas de ciencias de la salud que están accesibles en formato electrónico bajo la denominación de MEDLINE y EMBASE, respectivamente. Estas bases de datos referenciales pueden ser interrogadas mediante descriptores temáticos (*subject terms*) asignados por indexadores contratados por las empresas productoras. Los indexadores aplican terminología estandarizada a cada concepto específico de manera que los artículos que hagan referencia a un mismo concepto se pueden recuperar mediante un único descriptor temático. Usando los términos de búsqueda apropiados una sencilla estrategia de búsqueda puede identificar rápidamente artículos relativos al tema de interés. Este enfoque funciona bien si el objetivo es recuperar algún buen artículo sobre un tema o simplemente identificar un artículo específico. Sin embargo, cuando se realiza una revisión sistemática debe considerarse con saludable escepticismo la precisión con la que se aplican los descriptores a las referencias, dado que los autores quizá no describan bien sus métodos u objetivos; los indexadores no siempre son expertos en el área temática del artículo que están indexando y además están sujetos a los errores normales que se cometen cuando se realiza con rapidez un trabajo bastante tedioso. Por otro lado sólo están disponibles y son utilizados términos de indexación seleccionados, que pueden diferir de aquellos que utiliza

quien realiza la búsqueda. Ninguna de las bases de datos cubre todas las revistas y los artículos de las revistas que aparecen pueden haber estado incluidos de manera selectiva. Más adelante se discuten las estrategias para afrontar estas posibles limitaciones.

Se ha calculado que el solapamiento entre MEDLINE y EMBASE es aproximadamente de un 34% (Smith 1992). El volumen real de superposición de referencias –dependiendo del tema - oscila entre un 10% y un 75% (Kleijnen 1992, Smith 1992, Ramos-Remus 1994, Rovers 1993, Odaka.1992) Los estudios que comparan las búsquedas en las dos bases de datos han llegado, por regla general, a la conclusión que una búsqueda exhaustiva requiere que se interroguen ambas bases de datos. Estos estudios encontraron que aunque MEDLINE y EMBASE no identifican los mismos conjuntos de referencias, recuperan un número similar de referencias relevantes. Dado que puede esperarse que los resultados de las búsquedas en cada base de datos tengan una utilidad aproximadamente igual para la mayoría de los revisores, la decisión de cuál explorar primero puede quedar reducido, habitualmente, a un asunto económico, ya que MEDLINE es la opción más barata. Por ejemplo, *PubMed* (http://www.ncbi.nlm.nih.gov/PubMed/), proporciona acceso gratuito a través de internet a las citas bibliográficas de MEDLINE. También dispone de enlaces al texto completo de artículos alojados en las páginas web de los editores de las publicaciones . En el futuro se incluirán en PubMed otras bases de datos de la NLM, como AIDSLINE y *HealthStar .* La NLM está desarrollando una nueva base de datos a texto completo, que provisionalmente se llamará *Gateway,* la cual ofrecerá acceso en línea a resúmenes de congresos. Una característica particularmente útil de *PubMed* es que puede obtenerse una lista de "Artículos Relacionados" (*"Related Articles"*) a partir de cualquier registro identificado.

Los términos del lenguaje controlado no se han estandarizado entre las diferentes bases de datos, por lo que es necesario elaborar una estrategia de búsqueda a medida para cada una de ellas. Un método para identificar rápidamente los descriptores apropiados en una base de datos es obtener las palabras clave del texto de los artículos que hayan reunido los requisitos de inclusión así como los descriptores comunes que los indexadores han aplicado a todas ellas.

Durante los últimos años el Centro Cochrane del Reino Unido (*UK Cochrane Centre*) y el Centro Cochrane de Baltimore (*Baltimore Cochrane Center*) (ahora llamado *New England Cochrane Center, Providence Office (*NECC@P*)* han buscado en MEDLINE referencias desde 1966 hasta 1998 utilizando las dos fases de la estrategia de búsqueda altamente sensible de Cochrane (Dickersin 1994). Cientos de miles de registros han sido recuperados y revisados hasta la fecha Si, basándose en el título y el resumen, se considera que cumplen los criterios de la Colaboración Cochrane para ser definidos como ensayos controlados aleatorios o ensayos clínicos controlados se les asigna, en el campo Tipo de Publicación, de MEDLINE los descriptores *Randomized Controlled Trial o Controlled Clinical Trial* y se incluyen en CENTRAL/CCTR (con autorización de la NLM). (Ver Apéndice 5b. Para las definiciones de RCT y CCT según la NLM y Cochrane). De igual manera, en un proyecto actualmente en curso, el Centro Cochrane del Reino Unido (*UK Cochrane Centre)* está identificando registros en EMBASE, examinando sus títulos y resúmenes y enviándolos para que sean incluidos en CENTRAL/CCTR si se consideran apropiados (con autorización de *Elsevier)*.

5. 4. 3. SCISEARCH.

El *Science Citation Index* (Índice de Citas Científicas) y su versión electrónica Scisearch, identifica, a partir de un artículo publicado, los documentos en que éste ha

sido citado posteriormente. Identificar estudios para una revisión mediante Scisearch requiere que al menos exista un artículo clave sobre un tema y que aquellos que recojan investigaciones posteriores lo citen. Es una manera de buscar referencias posteriores a la publicación de un artículo importante.

5. 4. 4 Registros de ensayos clínicos

Existen registros que contienen referencias a ensayos en fase de planificación, en curso o finalizados, que han sido registrados prospectivamente. Esa puede ser una valiosa fuente de recuperación de estudios. La identificación de los ensayos al inicio de los mismos - y no en función de si han sido publicados o no- reduce el riesgo de sesgo de publicación que podría afectar a los resultados de una revisión. Se incluye, como apéndice, una lista de registros de ensayos clínicos (apéndice 5a).

5. 4. 5. Otras bases de datos bibliográficas

Dependiendo del tema de la revisión existen otras bases de datos bibliográficas que pueden resultar una fuente adicional de localización de ensayos para una revisión específica. Las bases de datos *The Gale Directory of Online*, *Portable* e Internet pueden utilizarse para localizar y evaluar la relevancia de otras bases de datos adicionales (http://www.dialog.com/).

Antes de emprender una nueva búsqueda siempre hay que consultar al Coordinador de la Editor Coordinador del Grupo de Revisión para determinar si ya se ha llevado a cabo la búsqueda con anterioridad.

5. 5. Búsqueda manual.

Para facilitar la identificación de todos los ensayos publicados, la Colaboración Cochrane ha organizado, a base de muchos esfuerzos, una amplia 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), o un ensayo clínico controlado (CCT).

La coordinación global de la búsqueda manual de la Colaboración en la literatura médica a nivel mundial es administrada por el NECC@P, que supervisa el registro prospectivo de todas las búsquedas manuales en la *Lista Maestra de revistas que son buscadas* (*Master List of Journals being Searched*. Hasta Junio de 2000, se habían realizado o estaban en curso búsquedas en más de 1700 revistas dentro de la Colaboración, las cuales se incluyen en la Lista Maestra (www.cochrane.org). La Lista Maestra permite registrar y controlar la progresión de la búsqueda para cada título de revista, y también ayuda a prevenir la duplicación de esfuerzos que podría surgir cuando grupos o individuos con especialidades que pueden solaparse revisan las revistas de su especialidad .

Los revisores que desean hacer búsquedas en revistas para identificar ensayos controlados potencialmente relevantes para su revisión, deben consultar primero a la Base Editorial de su Grupo. El Coordinador de Búsqueda de Ensayos del Grupo y/o el Coordinador del Grupo de Revisión determinaran si la revista ya ha sido revisada, y en caso de no ser así, el Coordinador puede registrar la búsqueda en la Lista Maestra y formar al revisor en búsqueda manual . Toda la correspondencia referente al inicio de la búsqueda en una revista, el progreso de la búsqueda en una revista, el estado de

la búsqueda, etc., tiene lugar entre el equipo del NECC@P y los Coordinadores de la Búsqueda de Ensayos y/o los Coordinadores del Grupo de Revisión.

5. 6. Desarrollo de una estrategia de búsqueda.

En el protocolo de la revisión debe describirse el proceso mediante el que se seleccionan los estudios para su inclusión en una revisión, es decir, debe explicitarse la estrategia de búsqueda. Estos procedimientos, cuanto más rigurosos son más reducen los riesgos de error y sesgo, a la vez que aumentan los recursos empleados y el tiempo invertido para preparar una revisión. La selección de los estudios que van a ser considerados para su inclusión en una revisión, es un proceso que incluye múltiples etapas. La primera etapa de revisión de los resultados de una búsqueda electrónica implica la evaluación detallada de los títulos y resúmenes para determinar si cada artículo reúne los requisitos predeterminados para ser escogido. Si al examinar la información disponible se puede determinar que un artículo no cumple los criterios de inclusión, el artículo puede rechazarse. Si, a criterio del revisor, el título o el resumen del artículo ofrecen dudas que impiden rechazarlo definitivamente, debería obtenerse el texto completo del artículo. La revisión del texto completo puede llevar a la exclusión del estudio porque no reúna los requisitos de inclusión. Si no se rechaza el artículo, este puede ser formalmente resumido como se describe en la sección 7. 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. Debido a la importancia que tienen estas decisiones, valdría la pena que más de un revisor examinara cada registro obtenido antes de que sea rechazado. 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.

Existe un rendimiento progresivamente menor en los esfuerzos de búsqueda, es decir que después de alcanzar un cierto nivel, cada unidad adicional de tiempo invertido en la búsqueda identifica menos referencias que sean relevantes para la revisión. En consecuencia, se llega a un punto donde el resultado de las búsquedas adicionales quizá ya no compense el esfuerzo requerido para identificar nuevas referencias. La decisión sobre cuánto invertir en el proceso de búsqueda depende del tema que aborde la revisión, el grado de desarrollo al que haya llegado el registro especializado del CRG y los recursos disponibles.

A menudo, los bibliotecarios o documentalistas que se han especializado en búsquedas electrónicas pueden ser de gran ayuda en el desarrollo y ejecución de las mismas. Estas personas pueden carecer de conocimientos sobre la materia de una revisión y estar entrenados para asegurar la precisión (es decir, llevan a cabo búsquedas que incluyen pocas citas no relevantes) mientras para una revisión sistemática la exhaustividad es más importante que la precisión. Aumentar la exhaustividad en una búsqueda implica reducir su precisión e identificar más artículos no relevantes. Aunque es importante que la búsqueda sea lo más completa posible, siempre es necesario mantener un equilibrio entre el nivel de exhaustividad y su precisión. El desarrollo de una estrategia de búsqueda es un proceso iterativo en el cual los términos que se utilizan son modificados según los resultados que se vayan obteniendo Por lo tanto, es importante trabajar estrechamente con un documentalista en el momento de decidir en qué bases de datos buscar y qué estrategias de búsqueda utilizar en cada una de ellas. Además, es útil suministrarle una o varias revisiones publicadas sobre el tema y sugerirle sinónimos de las palabras clave utilizadas. Idealmente, los revisores deben estar junto al documentalista cuando se ejecute la estrategia de búsqueda. Muchas veces hay un coste asociado a la búsqueda de cada base de datos y a los registros que son recuperados. Por lo tanto,

habitualmente es necesario considerar qué registros escoger en el mismo momento en que la búsqueda está llevándose a cabo. La fecha y los términos empleados en la interrogación a la base de datos, y el material identificado, deben ser registrados para cada búsqueda.

Un buen abordaje para desarrollar una estrategia de búsqueda exhaustiva es empezar utilizando términos múltiples que describan la condición o enfermedad de interés y combinarlos con el operador booleano "OR". Esto significa que se identificarán los artículos indexados que incluyan por lo menos uno de los términos utilizados. Los resultados pueden reducirse utilizando posteriormente el operador booleano “AND”. En este paso, se identifican los artículos que utilizan los términos que describen las intervenciones que se están evaluando, (agrupados con el operador “OR”) si también abordan la condición de estudio (es decir, si cumplen la intervención Y la condición). Aunque una pregunta de investigación puede referirse a poblaciones y situaciones específicas o a resultados concretos, estos conceptos a menudo no están bien indexados en las bases de datos electrónicas y por consiguiente son de difícil recuperación mediante búsqueda electrónica. Para reconocer qué ha sido correcta o incorrectamente indexado en una base de datos específica, es útil examinar los términos de lenguaje controlado aplicados a un grupo de estudios relevantes que hayan sido previamente identificados. En general, debe utilizarse una combinación de términos de lenguaje controlado y de palabras clave del texto (localizadas en el título o en el resumen).

Pueden utilizarse los términos metodológicos para hacer más precisa la recuperación si se están buscando tipos particulares de estudio (como ensayos aleatorios). La estrategia de búsqueda altamente sensible de Cochrane para MEDLINE ha sido diseñada para hacer búsquedas en la base de datos utilizando estos términos (apéndice 5c.). Esta estrategia puede combinarse (mediante el operador booleano "AND") con términos sobre la condición de interés, o pueden incorporarse después de identificar los artículos que abordan la condición Y las intervenciones de interés.

Mientras que MEDLINE y EMBASE incluyen citas desde 1966 y 1974, respectivamente, el Index Medicus y Excerpta Medica (versiones impresas de estas bases de datos) incluyen citas desde 1879 y 1948. La búsqueda manual en los primeros índices temáticos impresos puede ser útil si los estudios relevantes fueron publicados antes de la aparición de las bases de datos electrónicas. Por otro lado la NLM está construyendo una nueva base de datos electrónica, OLDMEDLINE, que incluye registros anteriores a 1966. Se encuentra disponible a través de *PubMed* (http://www.ncbi.nlm.nih.gov/PubMed/).

5. 7. Selección de estudios.

Todas las referencias de estudios identificados como potencialmente elegibles, deben ser evaluada posteriormente para ver si cumplen los criterios de inclusión para la revisión. Los revisores deben decidir:

- Cuando se necesitará que más de un revisor evalúe 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 conocerán 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 en una revisión a menudo implican un juicio. Para ayudar a asegurar su reproductibilidad es aconsejable que más de un revisor aplique los criterios de inclusión a todos los artículos evaluados y potencialmente relevantes. Sin embargo, el número necesario de juicios varía de revisión en revisión. Algunos revisores quizá decidan que las opiniones acerca de los estudios sean muy simples. En estos casos el uso de múltiples revisores para evaluar la relevancia quizá resulte trivial. Sea como fuere, el número de personas que evalúan la relevancia de cada estudio debe informarse en la sección de Métodos de la revisión (si no se ha realizado una descripción de los métodos usados por todos los revisores en un CRG específico).

Los especialistas en un área específica con frecuencia tienen opiniones formadas previamente que pueden sesgar sus evaluaciones, tanto sobre la relevancia como sobre la validez de los artículos (Cooper 1989, Oxman 1993b). Por lo tanto, aunque es importante que al menos un revisor sea versado en el tema objeto de revisión, quizá sea una ventaja disponer de un segundo revisor que no sea un experto en esa área.

Algunos revisores pueden decidir que las evaluaciones sobre la relevancia de los artículos deben ser hechas por personas que, en el momento de preparar las copias de los artículos, desconozcan la revista de la cual procede el artículo, los autores, la institución, magnitud y sentido de los resultados, (Berlin 1997a, Berlin 1997b). Sin embargo, este proceso requiere mucho tiempo y puede no ser razonable llevarlo a cabo dados los recursos que requiere y la incertidumbre de que proteja de los sesgos (Berlin 1997b).

Las discrepancias acerca de si un ensayo debe incluirse o no en la revisión pueden resolverse generalmente mediante discusión. A menudo, la causa del desacuerdo es un simple descuido por parte de uno de los revisores. Cuando la discrepancia se debe a una diferencia de interpretación el tema debe ser resuelto por consenso. Cuando, por falta de información no sea posible resolver las diferencias acerca de incluir o no un estudio - éste quedará a la espera de evaluación mientras no se obtengan los datos adicionales.

Para la mayoría de las revisiones puede ser útil realizar una prueba piloto de aplicación de los criterios de inclusión a una muestra de artículos (sobre unos diez o doce, incluidos aquellos que se supone que son definitivamente elegibles, definitivamente no elegibles y dudosos). La prueba piloto puede usarse para pulir y clarificar los criterios de inclusión, formar a las personas que los aplicarán y asegurarse de que los criterios puedan ser aplicados sistemáticamente por más de una persona.

Un método para determinar qué estudios se identifican como excluidos en la revisión consiste en elaborar una lista con los estudios de los que cabría esperar que el lector cuestione por qué no fueron incluidos. Esto incluye todos los estudios que aparentemente cumplen los criterios de selección pero que se excluyeron, así como los que no cumplieron todos los criterios de inclusión pero que son bastante conocidos en el área general en que se desarrolla la revisión y que probablemente pueden ser relevantes para algunos lectores. Al elaborar con ellos una lista de estudios excluidos y proporcionar las principales razones para su exclusión, el revisor puede demostrar que han sido tenidos en cuenta.

5. 8. Cómo documentar una estrategia de búsqueda.

La estrategia de búsqueda debe describirse con suficiente detalle en la revisión como para que el proceso pueda reproducirse. La documentación debe incluir las principales fuentes de información utilizadas para identificar los estudios, la estrategia de búsqueda para cada base de datos consultada y la fecha de la búsqueda, las limitaciones temporales (por ejemplo, de 1966 a 1996) y las limitaciones de idioma o de otro tipo, si las hubiese (aunque deben evitarse en la medida de lo posible). Como señalábamos anteriormente, debe informarse cómo y cuándo fue examinado por última vez el registro especializado del CRG para identificar estudios relevantes. La fecha en que se realizó la interrogación a la base de datos debe incluirse dentro de la sección del resumen correspondiente a la estrategia de búsqueda. Pueden encontrarse ejemplos de las descripciones de las estrategias de búsqueda en la CDSR. Hay una variación muy amplia en lo extensiva que hayan hecho la búsqueda los revisores más allá de las referencias incluidas en el registro especializado del CRG así como en el detalle con el que se describe la estrategia de búsqueda. Aunque lo ideal sea una estrategia de búsqueda suficientemente desarrollada para hacer innecesarias las búsquedas adicionales, la mayoría de los grupos de revisión no ha llegado a este nivel de desarrollo.

Dado que las revisiones Cochrane se actualizan constantemente, las búsquedas continúan en curso, y tanto las estrategias de búsqueda como el informe de las mismas pueden seguir mejorándose. Para facilitar la documentación de las estrategias de búsqueda se sugiere utilizar la lista de las instrucciones de búsqueda (es decir, “cortar” y “pegar” la estrategia en la sección correspondiente de la revisión en vez de introducir nuevamente la información) para describir lo que se hizo, junto con el título de la base de datos, la fecha en que se ejecutó la búsqueda y los años cubiertos por la misma. Por ejemplo:

MEDLINE en Silver Platter (version 2.0 de 1995) fue consultada en Enero de 1996, para el periodo entre 1966 hasta 1999, usando la siguiente estrategia de búsqueda (seguido de una impresión directa de la estrategia de búsqueda).

Las estrategias de búsqueda usadas para desarrollar registros especializados de ensayos de los CRG se describen en el módulo del CRG, y no hace falta que se documenten en el texto de las revisiones Cochrane.

5. 9. Seguimiento de los estudios identificados.

Cuando se recogen referencias de múltiples fuentes o cuando se ejecutan estrategias de búsqueda múltiples en la misma base de datos es habitual duplicar la misma referencia más de una vez. En lo posible, estos duplicados deben eliminarse antes de evaluar los informes. Así mismo los servicios de préstamo interbibliotecario y de obtención de documentos pueden perder las solicitudes, o quizá no lleguen a entregar un artículo hasta meses después de que se haya solicitado. Un sistema de gestión de referencias como *Procite,* que facilita el trabajo de identificación de títulos duplicados y ayuda en el seguimiento del proceso de recuperación de cada artículo. También es buena idea documentar el motivo de exclusión de cada artículo aunque al final haya sido rechazado. Cuando se efectúan búsquedas en múltiples bases de datos puede ser útil hacer un seguimiento de que referencias se identificaron en cada una de ellas ya que puede guiar acerca de qué bases de datos explorar rutinariamente para mantener el registro especializado del CRG y si fuera conveniente, para las revisiones individuales.

Los programas informáticos especialmente diseñados para el manejo de referencias como ProCite y Reference Manager son útiles y relativamente fáciles de usar. Los

paquetes generales de bases de datos como Access y FoxPro ofrecen alta capacidad de consulta y fácil adaptación, pero requieren al menos una programación mínima y suficiente destreza en el diseño de bases de datos.

Los CRG están usando varios programas diferentes para administrar sus registros especializados pero ProCite es el programa más utilizado y el único para el cual se dispone ampliamente en los distintos equipos editoriales de los CRGs. También es la base de datos preferida para enviar los ensayos controlados y los registros especializados al CENTRAL/CCTR. El programa MeerKat ha sido desarrollado por el Centro Cochrane del Reino Unido (*UK Cochrane Centre)* en asociación con *Update Software* para hacer frente a las necesidades de los CRG en cuanto al manejo de sus registros especializados. Si finalmente es adoptado facilitará la tarea del manejo de las referencias dentro de los CRG.

5. 10. Contribución a los registros especializados

Para asegurar la viabilidad a largo plazo de los CRG y de la Colaboración Cochrane es importante coordinar los esfuerzos para desarrollar y mantener los registros especializados de manera eficaz y eficiente. Es necesario facilitar el acceso de los revisores a los estudios, ayudando a que las revisiones sean exhaustivas, actualizadas e imparciales. También es necesario protegerse contra la duplicación de esfuerzos y minimizar la carga que recae en los revisores para llevar a cabo sus propias búsquedas.

Para obtener el máximo beneficio y reducir al mínimo los esfuerzos desaprovechados, es importante que los revisores trabajen estrechamente con los coordinadores de búsquedas de ensayos y con los coordinadores de los grupos de revisión en el desarrollo y ejecución de las búsquedas que ellos realizan. Específicamente, podemos mencionar:

- El desarrollo de una estrategia de búsqueda debe empezar con la consulta del registro especializado del CRG y con la familiarización de los métodos usados para el desarrollo y mantenimiento del registro. El equipo editorial de un CRG debe poder orientar y ayudar en el desarrollo y puesta en práctica de una estrategia de búsqueda.
- Los estudios identificados por los revisores a partir de otras fuentes que están dentro del ámbito del CRG, deben incorporarse al registro especializado. Cada CRG debe establecer los mecanismos apropiados para facilitar la contribución de estudios desde los revisores hasta su registro especializado. El registro debe enviarse para que sea incluido en el CENTRAL/CCTR.
- Los revisores quizá deseen contribuir de manera específica al desarrollo y mantenimiento del registro especializado más allá de lo estrictamente necesario para una revisión en particular (como por ejemplo, búsquedas manuales de revistas o de presentaciones a congresos). Cualquier búsqueda manual que se emprenda debe consultarse con el equipo editorial del CRG.
- La comunicación personal con expertos en un campo puede contribuir al desarrollo del registro especializado del CRG así como a la de una revisión específica. Puede ser útil especificar planes para contactar con expertos contando con la ayuda del equipo editorial del CRG para coordinar estos esfuerzos.
- Asimismo, los revisores pueden ayudar a difundir y revisar las estrategias de búsqueda usadas para mantener un registro especializado basándose en la experiencia que tienen al realizar búsquedas para sus propias revisiones.

5.11 Evidencia sobre efectos adversos

La primera fuente para buscar información acerca de efectos adversos son los estudios incluidos en la revisión. Aquellos estudios que han sido excluidos también pueden proporcionar información útil, dependiendo de las razones por las cuales fueron excluidos.

Si es de interés los efectos adversos de los medicamentos existen numerosas fuentes de información, tales como *Current Problems* producida por la *UK Medicines Control Agency* (http://www.open.gov.uk/mca/mcahome.htm), *MedWatch* producida por la *US Food and Drug Administration*, y el *Australian Adverse Drug Reactions Bulletin* (http://www.health.gov.au/). Las agencias de regulación y los fabricantes de medicamentos también pueden proporcionar algún tipo de información. Para todos los tipos de intervención también puede buscarse información acerca de efectos adversos en otros tipos de estudios diferentes a los considerados dentro de la revisión (como estudios de cohortes, de casos y controles, ensayos no controlados, series de casos e informes de casos). Sin embargo, todos estos estudios e informes tienden a estar más sesgados que los ensayos aleatorios, y deben ser interpretados con cautela.

5.12 Referencias

Begg 1988 Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. J R Statist Soc A 1988; 151: 445-63.

Begg 1989 Begg CB, Berlin JA. Publication bias and dissemination of clinical research. J Natl Cancer Inst 1989; 81: 107-15.

Berlin 1997a. Berlin JA, Miles CG, Crigliano MD. Does blinding of reasers affect the results of meta-analyses? Results of a randomized trial. Online J Curr Clin Trials 1997

Berlin 1997b. Berlin JA. Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Bliding Study Group [letter]. Lancet 1997; 350: 185-6.

Boynton 1998. Boynton J, Glanville J, McDaid D, Lefebvre. Identifying systematic reviews in MEDLINE: developing an objective approach to search strategy design. Journal of Information Science 1998;24:137-57

Cooper 1989. Cooper HM, Ribble RG. Influences on the outcome of literature searches for integrative research reviews. Knowledge 1989; 10:179-201.

Dickersin 1987. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias and clinical trials. Controlled Clin Trials 1987; 8:343-53.

Dickersin 1992 Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. JAMA 1992; 263: 374-8.

Dickersin 1993. Dickersin K, Min YI. NIH clinical trials and publication bias. Online J Curr Clin Trials [serial online] 1993; Doc No 50.

Dickersin 1994. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. BMJ 1994; 309:1286-91.

Easterbrook 1991. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991; 337:867-72.

Egger 1997. Zellweger Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. Lancet 1997; 350:326-9.

Gotzsche 1987. Gotzsche PC. Reference bias in reports of drug trials. BMJ 1987; 295:654-6

Gregoire 1995. Gregoire G, Derderian F, LeLorier J. Selecting the language of the publications included in a meta-analysis: is there a tower of Babel bias? J Clin Epidemiol 1995; 48.159-63.

Hetherington 1989. Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. Pediatrics 1989; 84:374-80.

Kleijnen 1992. Kleijnen J, Knipschild P. The comprehensiveness of Medline and Embase computer searches. Searches for controlled trials of homeopathy, ascorbic acid for common cold and ginkgo biloba for cerebral insufficiency and intermittent claudication. Pharm Weekbl Sci 1992; 14:316-20.

Moher 1996a. Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. Lancet 1996; 347:363-6.

Odaka 1992. Odaka T, Nakayama A, Akazawa K, Sakamoto M, Kinukawa N, Kamakura T, et al. The effect of a multiple literature database search a numerical evaluation in the domain of Japanese life science. J Med Syst 1992; 16:77-81.

Oxman 1993b. Oxman AD, Guyatt GH. The science of reviewing research. Ann NY Acad Sci 1993; 703:125-33.

Ramos-Remus 1994. Ramos-Remus C, Suarez-Almazor M, Dorgan M, Gomez-Vargas A, Russell AS. Performance of online biomedical databases in rheumatology. J Rheumatol 1994; 21(10):1912-21.

Ravnskov 1992. Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. BMJ 1992; 305:15-9.

Rovers 1993. Rovers JP, Janosik JE, Souney PF. Crossover comparison of drug information online database vendors: Dialog and MEDLARS. Annals of Pharmacotherapy 1993; 27(5):634-9.

Scherer 1994. Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. JAMA 1994; 272: 158-62.

Simes 1986. Simes RJ. Publication bias: the case for an international registry of clinical trials. J Clin Oncol 1986; 4:1529-41.

Simes 1987. Simes RJ. Confronting publication bias: a cohort design for meta-analysis. Stat Med 1987; 6:11-29.

Smith 1992. Smith BJ, Darzins PJ, Quinn M, Heller RF. Modern methods of searching the medical literature. Med J Aust 1992; 157:603-11.

6. EVALUACIÓN DE LA CALIDAD DE LOS ESTUDIOS.

La evaluación de la calidad de los estudios individuales que se incluyen en las revisiones sistemáticas es necesaria para limitar los sesgos, formarse una idea más precisa de las potenciales comparaciones que puedan establecerse y guiar la interpretación de resultados. Los factores que justifican una evaluación son aquellos relacionados con la aplicabilidad de los resultados, la validez de los estudios individuales y ciertas características del diseño que afectan a la interpretación de los resultados.

La aplicabilidad - también llamada validez externa o generalización - está relacionada con la definición de los componentes clave de las preguntas bien formuladas que han sido esbozados en la sección 4. Específicamente, el hecho de que los resultados de una revisión sean aplicables a una población específica, a una estrategia de intervención o resultado depende de los estudios seleccionados para la revisión y, a su vez, de cómo los estudios y los propios revisores definen a las personas, las intervenciones y los resultados de interés.

La interpretación de los resultados depende de la validez de los estudios incluidos y de otras características. Por ejemplo, una revisión quizá resuma veinte ensayos clínicos válidos que evalúan los efectos de los tratamientos anti-isquémicos sobre los síntomas del dolor torácico en adultos con infarto de miocardio previo. Sin embargo, los ensayos clínicos quizá examinen diferentes preparados y dosis de los agentes anti-isquémicos y es posible que tengan duraciones variables. Estos últimos aspectos afectarían a la interpretación aunque quizá no estén relacionados directamente con la validez interna de los ensayos. En la próxima sección (sección 7) se darán ejemplos de cómo pueden extraerse los datos relacionados con la aplicabilidad y el diseño del estudio que puedan afectar la interpretación de los resultados. El resto de esta sección se centrará en evaluar la validez de los estudios individuales incluidos en una revisión sistemática. Dado que la mayoría de las revisiones Cochrane se centran en ensayos clínicos aleatorios, nos centraremos en cómo evaluar la validez de dichos estudios.

6. 1. Validez

En el contexto de una revisión sistemática, 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 (Moher 1995). Un concepto importante que no debe ser confundido con la validez es la precisión. La precisión es una medida de la probabilidad de que se observe un efecto por puro azar, lo que conduce a una conclusión falsamente positiva (error aleatorio). Se refleja en el intervalo de confianza alrededor del efecto estimado en 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). A los resultados más precisos se les da más peso.

En una revisión sistemática, las diferencias en la validez de los estudios incluidos pueden explicar la variabilidad de sus resultados. Los estudios más rigurosos tienen mayor probabilidad de dar resultados más próximos a la “verdad”. Si los estudios menos rigurosos están sesgados hacia la sobreestimación de la eficacia de una intervención, el análisis cuantitativo de los resultados de estudios de diferente grado de validez puede dar lugar a conclusiones "positivas falsas" (concluyendo erróneamente que una intervención es eficaz). También se puede llegar a conclusiones "negativas falsas" (concluyendo erróneamente que no tienen ningún

efecto) si los estudios menos rigurosos están sesgados de manera que subestiman el efecto de una intervención (Detsky 1992).

Aunque no haya diferencias ni en la validez ni en los resultados de los estudios incluidos en una revisión, es importante evaluar críticamente todos los estudios de manera sistemática. Por ejemplo, podría ocurrir que los resultados fueran consistentes en todos los estudios evaluados pero que los estudios fueran imperfectos. En este caso, las conclusiones de la revisión no serían tan sólidas como si una serie de estudios rigurosos produjeran resultados consistentes acerca del efecto de una intervención.

6. 2. Tipos de sesgo en los ensayos clínicos sobre intervenciones sanitarias.

En los estudios sobre los efectos de la atención sanitaria hay cuatro tipos de sesgo o errores sistemáticos: el sesgo de selección, el sesgo de realización (*performance bias*), el sesgo de desgaste o pérdida (*attrition bias*) y el sesgo de detección (ver figura abajo). Lamentablemente, no hay evidencia empírica sólida sobre la relación entre los resultados de los estudios y los criterios usados para evaluar el riesgo de estos sesgos (Moher 1995, Moher 1996b). Sin embargo, hay fundamentos para creer que esta relación existe y por lo tanto, es razonable tener en cuenta estos cuatro sesgos potenciales al momento de evaluar los estudios para una revisión (Feinstein 1985).

• 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)	Medidas de resultado	Medidas de resultado

6. 3. Sesgo de selección

Uno de los factores más importantes que puede sesgar y distorsionar las comparaciones entre tratamientos está relacionado con el proceso de obtención de los grupos a comparar (Kunz 1998). En el diseño de ensayos clínicos es muy importante utilizar un método apropiado para prevenir que se conozca de antemano la asignación de los tratamientos.

Mientras se evalúa la elegibilidad potencial de un participante en un ensayo clínico, tanto las personas que reclutan a los participantes como los mismos participantes, deben desconocer el tratamiento que corresponde a la próxima asignación. Una vez el participante resulta elegido y se conoce la asignación, no deben poder alterar la asignación o la decisión acerca de la elegibilidad.

Lo ideal es que el proceso de asignación no pueda ser influido por las personas que hacen la asignación. Para alcanzar este objetivo debería instaurarse un programa de asignación aleatoria implantado por alguien no implicado en el reclutamiento. Por ejemplo, alguien ubicado en una oficina central del ensayo o en el servicio que entrega los medicamentos. Si no puede organizarse tal aleatorización central, entonces es necesario tomar otras precauciones para evitar la manipulación del proceso de asignación por las personas que participan en el reclutamiento.

El proceso de ocultar la asignación (*allocation concealment*) hasta que se asigna el tratamiento se ha denominado también “cegamiento de la aleatorización" (*randomization blinding*) (Chalmers 1983). Esta expresión no distingue claramente el

proceso de ocultación (o encubrimiento de la asignación) del proceso de cegamiento de los pacientes, de los proveedores sanitarios, de quienes evalúan los resultados o de los que analizan los datos. Además, no es satisfactoria por tres razones. En primer lugar, la razón para ocultar el proceso de asignación es eliminar el sesgo de selección. En cambio, el cegamiento (usado después de la asignación de la intervención) reduce los sesgos de realización y de detección. En segundo lugar, desde un punto de vista práctico, el encubrimiento de la asignación hasta el momento mismo de la asignación siempre es posible, independientemente del tema de estudio, pero el cegamiento después de la asignación quizá sea imposible, como sucede en los ensayos clínicos que comparan un tratamiento quirúrgico con un tratamiento médico. Y tercero, el control del sesgo de selección es importante para el ensayo clínico en su totalidad, y por lo tanto, para todos los resultados que están siendo comparados.

Por otro lado, a menudo, el sesgo de detección se pueda controlar para algunos de los resultados de un estudio pero no para otros. Por lo tanto, los procesos de ocultación hasta la asignación y de cegamiento posterior a la asignación afrontan diferentes tipos de sesgo, su forma de aplicación es intrínsecamente distinta y se refieren a diferentes componentes de un estudio. Para distinguir claramente estos propósitos y formas de "cegar", nos referiremos al proceso de "encubrir" las asignaciones como "ocultación de asignaciones" y reservaremos el término "cegar" para las medidas tomadas para reducir el sesgo una vez que la intervención ya ha sido asignada.

La investigación empírica ha mostrado que la falta de ocultación de las asignaciones se asocia a la presencia de sesgos (Chalmers 1983, Schulz 1995, Moher 1998a). En realidad, se ha visto que la ocultación es más importante en la prevención del sesgo que otros aspectos de la asignación, como por ejemplo el proceso de generación de la secuencia de asignación (programa de ordenador, tabla de números aleatorios, alternancia).

Por lo tanto, los estudios pueden evaluarse teniendo en cuenta el método de ocultar las asignaciones. Debería presentarse información que proporcione credibilidad de que las asignaciones no se conocían por lo menos hasta antes de la asignación. El método de asignación de los participantes a las intervenciones debe ser suficientemente sólido contra la influencia de pacientes y clínicos, y su descripción debe ser clara. A continuación se presentan algunas propuestas para asegurar mecanismos de ocultación adecuados:

- Asignación centralizada (por ejemplo, asignación de los pacientes por una oficina central sin conocimiento de las características de los sujetos) o aleatorización controlada por un servicio de farmacia.
- Utilización de envases idénticos numerados de antemano y administrados secuencialmente a los participantes.
- Utilización de un sistema informatizado local combinado con asignaciones que se guardan en archivos electrónicos ilegibles y cerrados, a los cuales puede accederse únicamente después de que se han introducido las características de un participante reclutado.
- Sucesión de sobres numerados, sellados y opacos.

Otros abordajes pueden incluir estrategias similares a las ya enumeradas anteriormente, junto con la garantía de que la persona encargada de la asignación no sea quien lo administra. Algunos mecanismos pueden ser innovadores y no adecuarse a ninguno de los enfoques anteriores, pero pueden igualmente proporcionar un método de ocultación adecuado.

Los métodos de ocultación de las asignaciones claramente inadecuados incluyen: alternancia, el empleo de números de historia clínica, la fecha de nacimiento o del día de la semana y cualquier procedimiento que sea completamente transparente antes de la asignación, como puede ser una lista abierta de números aleatorios. Cuando los estudios no informen del método de ocultación, su rigor es dudoso. Algunos ejemplos incluyen la mera declaración de que se usó una lista o tabla, la sola especificación de que se usaron sobres sellados o la información de un método de ocultación que parece adecuado, combinada con otra información que produzca suspicacia en el revisor. Cuando los revisores introduzcan los estudios en el programa *RevMan* éste les preguntará si el sistema de ocultación de las asignaciones fue adecuado (A), poco claro (B), inadecuado (C), o si no se usó la ocultación de las asignaciones (D), como criterio para evaluar la validez.

6. 4. Sesgo de realización

Con “sesgo de realización” (*performance bias*) se hace referencia a las diferencias sistemáticas en la atención sanitaria proporcionada a los participantes en los grupos de comparación, a parte de la intervención bajo estudio. 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 están asignados quienes la reciben. Hay estudios que indican que tales sistemas de cegamiento son importantes en la prevención del sesgo (Karlowski 1975, Colditz 1989, Schulz 1995).

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 (CCSG 1978, Sackett 1979b). Además, hay evidencia de que los participantes que son conscientes de su grupo de asignación comunican más síntomas, y por lo tanto, se obtienen resultados sesgados (Karlowski 1975). Por estas razones, los revisores deberían considerar el uso del "cegamiento" como un criterio de validez. Esto puede hacerse con las siguientes preguntas: ¿Quiénes recibieron atención, desconocían la intervención que les había sido asignada? ¿Aquellos que proporcionaron la atención, desconocían la intervención asignada?

A menudo se agrega una tercera pregunta relacionada con el cegamiento y el sesgo de detección: ¿Los responsables de evaluar los resultados, desconocían la intervención asignada? Esto tiene relación con el sesgo de detección, como se menciona más adelante.

Los revisores que trabajan en áreas donde el cegamiento es importante deberían 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. En este caso, el Grupo Colaborador de Revisión debería consensuar un enfoque estándar para evaluar el cegamiento (Chalmers 1989, Schulz 1995, Jadad 1996, Moher 1996b).

6. 5. Sesgo de desgaste

El sesgo de desgaste (*attrition bias*) se refiere a las diferencias sistemáticas entre los grupos de comparación en cuanto a la pérdida de participantes durante el estudio. Ha sido también denominado “sesgo de exclusión” pero aquí lo denominamos sesgo de desgaste o de pérdida para evitar la confusión con los criterios de exclusión y de inclusión previos al reclutamiento de los participantes. Cuando la información sobre cómo se tratan las pérdidas de participantes (por abandonos, bajas, desviaciones de

protocolo, etc.) es insuficiente, los revisores deben tener mucha precaución con los cálculos del seguimiento.

La manera de tratar las pérdidas puede sesgar los resultados, y la información incompleta sobre cómo se tratan las pérdidas puede ocultar este problema. Hasta ahora no se ha encontrado una relación consistente entre las pérdidas posteriores a la asignación y el sesgo (Schulz 1995). Por lo tanto, los revisores deben ser prudentes al utilizar la información de seguimiento como un criterio de validez, en particular cuando ésta se detalla de manera implícita. Sin embargo, ésta es una recomendación general y puede no ser aplicable a ciertas áreas temáticas que tienen buena calidad de la información o en las que es posible obtener de los investigadores la información que falte.

6. 6. Sesgo de detección

El sesgo de detección se refiere a las diferencias sistemáticas entre los grupos de comparación en la evaluación de resultados. Los ensayos clínicos que ciegan la asignación de las intervenciones a quienes evalúan resultados son, lógicamente, menos propensos a estar sesgados que los ensayos clínicos que no lo hacen. El cegamiento es particularmente importante en aquellas investigaciones que utilizan medidas subjetivas de resultado, como el dolor (Karlowski 1975, Colditz 1989, Schulz 1995). Sin embargo, al menos dos estudios empíricos no han podido demostrar una relación entre el cegamiento de la evaluación de resultados y los resultados del estudio. Esto puede ser debido a insuficiencias en el reporte de los estudios (Reitman 1988).

El sesgo debido a un reporte selectivo de resultados es diferente del sesgo en la evaluación de resultados. Esta fuente de sesgo puede ser importante en áreas donde se usan múltiples medidas de resultado, como en las evaluaciones de tratamientos para la artritis reumatoidea (Gotzsche 1989). Por consiguiente, los revisores deberían considerar como indicadores de validez la especificación *a priori* de las medidas de resultado y de los análisis por parte de los investigadores. Además, la presentación selectiva de resultados específicos debería utilizarse para promover la presentación de mejores informes y los esfuerzos de los revisores en la obtención de los datos que faltan.

6. 7. Métodos para resumir la validez de los estudios

6. 7. 1. 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 individuales y al mismo tiempo, usar los criterios individuales, como el grado de ocultación de las asignaciones, en los análisis de sensibilidad (ver sección 8.10). Sin embargo, si se han utilizado varios criterios explícitos para evaluar la validez, es aconsejable resumirlos para obtener una evaluación general de la validez de los resultados de cada estudio. Un enfoque sencillo para hacerlo es usar tres categorías:

Riesgo de sesgo	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 en los resultados	Al menos un criterio no se cumple

Las relaciones sugeridas en la tabla serían apropiadas si sólo se usara un número reducido de criterios de evaluación y si todos los criterios trataran de aspectos importantes para la validez de los resultados del estudio. En general, y cuando sea posible, los revisores deberían obtener información adicional de los autores cuando no queda claro si se cumplió o no un criterio.

6. 7. 2. Escalas y listas de “calidad"

David Moher y sus colaboradores han identificado 25 escalas y 9 listas o *checklists* que se han utilizado para evaluar la validez y la "calidad" de los ensayos clínicos controlados y aleatorizados (Moher 1995, Moher 1996b). Estas escalas y listas incluyen entre 3 y 57 elementos y para completarlas se necesita entre 10 y 45 minutos. Casi todos los elementos se basan en criterios sugeridos o "generalmente aceptados" que se mencionan en los libros de texto sobre ensayos clínicos.

Muchos de los instrumentos de medida de la calidad tienden a confundir la calidad de los informes con la validez del diseño y la realización de un ensayo clínico. Además, la calificación se basa en si algo se informó (por ejemplo, cómo se asignaron los participantes) en lugar de si se hizo apropiadamente en el estudio. Muchos contienen también elementos que no están directamente relacionados con la validez, tales como si se hizo un cálculo de la potencia (un elemento que está más relacionado con la precisión de los resultados) o si los criterios de inclusión y exclusión se describieron claramente (un elemento que se relaciona más con la aplicabilidad que con la validez).

Dado que no hay ningún *"gold standard"* para determinar la validez "real" de un ensayo clínico, la posibilidad de validar cualquier sistema de medida de la validez es limitada. Aunque es posible aplicar los principios básicos de medición al desarrollo de una escala de evaluación de la validez de los ensayos clínicos aleatorizados, la relación entre tal sistema de puntuación y el grado en que un estudio está libre de sesgo no es obvia. Ninguna de las escalas actualmente disponibles para medir validez o "calidad" de los ensayos clínicos puede recomendarse sin reservas. Si los revisores o los Grupos Colaboradores de Revisión deciden usar una escala de este tipo, deben tener precaución.

La mayoría de las escalas disponibles para evaluar la validez de los ensayos clínicos controlados y aleatorizados proporcionan una puntuación resumen mediante la suma de las puntuaciones (con o sin pesos diferenciales) de cada elemento. Aunque este enfoque puede ser atractivo por su sencillez, no se sustenta en estudios empíricos (Emerson 1990, Schulz 1995).

De manera notable, lleva más tiempo completar las escalas con múltiples elementos y sistemas de puntuación complejos que utilizando métodos sencillos. No se ha

demostrado que las escalas más complejas proporcionen evaluaciones de la validez que sean más fiables (Jüni 1999). Además, pueden comportar un mayor riesgo de confundir la calidad de los informes con la validez del estudio. Con mayor probabilidad tienden a incluir criterios que no miden directamente la validez interna y tienden a ser menos transparentes a los usuarios de la revisión. Por estas razones, para evaluar la validez es preferible usar métodos sencillos que puedan informarse claramente (por ejemplo, cuántos puntos obtuvo cada ensayo clínico en cada criterio).

6. 8. Sesgo en los estudios no experimentales

El Grupo de Métodos de Estudios No Aleatorios está preparando unas guías acerca del uso de estudios no aleatorios en las revisiones Cochrane (apéndice 6). Mientras tanto, esta sección describe algunos aspectos que deben tenerse en cuenta al momento de evaluar la validez de los estudios no aleatorios. El motivo principal de que las revisiones Cochrane se centren en ensayos clínicos controlados y aleatorizados es que la aleatorización es el único medio de asignación que controla tanto los factores de confusión (*confounders*) desconocidos y no medibles como aquellos que se conocen y pueden medirse. Las diferencias entre los grupos de comparación en cuanto al pronóstico, sensibilidad (*responsiveness*) al tratamiento, o la exposición a otros factores que afectan a los resultados pueden distorsionar la magnitud aparente de los efectos de la intervención de interés. En los estudios observacionales, como los estudios de casos y controles y los estudios 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.

Lamentablemente, rara vez puede darse por sentado que se conozcan todos los factores importantes relacionados con el pronóstico y con la sensibilidad o capacidad de respuesta al tratamiento, y en caso de conocerlos pueden surgir dificultades a la hora de medirlos o de tenerlos en cuenta en el análisis. Estas preocupaciones están avaladas por la evidencia empírica (Kunz 1995).

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, el sesgo de selección, por término medio, tiende a hacer que los efectos del tratamiento parezcan más grandes de lo que son y el tamaño de estas distorsiones puede ser tan grande, o incluso mayor, que el tamaño de los efectos que están midiéndose (Kunz 1995).

A pesar de estas inquietudes, a veces hay razones de peso para confiar en los estudios observacionales a fin de obtener información acerca de los efectos de las intervenciones de la atención sanitaria e incluir tales estudios en las revisiones Cochrane. Por ejemplo, estudios observacionales bien diseñados han proporcionado información útil sobre los efectos de las intervenciones como es el uso obligatorio de los cascos por los motociclistas, el *screening* (cribaje, tamizaje) de cáncer de cuello uterino, la difusión de guías de práctica clínica para cambiar la práctica profesional y los efectos adversos poco frecuentes de la medicación.

Se han sugerido diversos criterios a fin de evaluar críticamente la validez de los estudios observacionales (Horwitz 1979, Feinstein 1982, Levine 1994, Bero 1999). En general, los mismos cuatro tipos de sesgo comentados anteriormente pueden aplicarse a otros tipos de estudios comparativos, como se ilustra a continuación:

Los aspectos relacionados con el sesgo de desgaste son similares en el caso de los ensayos clínicos aleatorizados, los estudios de cohortes y los estudios de casos y controles, y se relacionan con el grado en que todos los participantes incluidos en un estudio son tenidos en cuenta adecuadamente en los resultados. Los aspectos relacionados 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 aleatorizados 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, los revisores 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 del grupo de comparación a otros factores que podrían afectar los resultados.

Además de las consideraciones sobre el método de cegamiento, similares a las de los ensayos clínicos aleatorizados, es importante evaluar si la exposición se midió de manera similar y no sesgada en los grupos que se comparaban. Así, por ejemplo, en los estudios de cohortes y de casos y controles sobre los efectos del tratamiento hormonal sustitutivo en mujeres post-menopáusicas, además de los aspectos relacionados con el sesgo producido por factores de confusión, los investigadores y los revisores deben asegurar que la administración de hormonas se midió de manera no sesgada.

En resumen, es necesario conocer muy bien el tema al evaluar la validez de los estudios observacionales. También se necesita experiencia cuando se evalúa la validez de los ensayos clínicos aleatorizados, 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 y por ello, se necesita tener precaución.

6. 9. Aplicación de los criterios de evaluación de la calidad

Cuando se hace una evaluación de los estudios deben tomarse varias decisiones básicas, similares a las que se hacen en el proceso de selección de estudios (sección 5.7). Un aspecto importante es el número de revisores. ¿Debe haber uno o más de uno? ¿Cuántos se necesitan y cuándo los hay en exceso? ¿Deben revisar los mismos artículos a fin de maximizar la fiabilidad o revisarán subconjuntos de estudios mutuamente excluyentes a fin de minimizar el volumen de trabajo? Al mismo tiempo, debe tenerse en cuenta la formación de los distintos revisores y si éstos deben tener experiencia previa en diseño de estudios y en evaluación crítica de los mismos.

La realización de revisiones sistemáticas con múltiples revisores presenta las dos caras de la moneda. Por un lado, puede limitar los sesgos, minimizar los errores y mejorar la fiabilidad de los resultados, pero cuando hay más de un revisor puede haber discrepancias entre los revisores. Cuando van a participar múltiples revisores, debe haber un procedimiento o regla de decisión explícita establecida *a priori* para identificar y resolver los desacuerdos. Como regla general, se recomienda que al menos dos revisores evalúen la información que implica una interpretación subjetiva y la

información que pueda afectar a la interpretación de resultados (por ejemplo, datos de resultado). La sección 7 describe los métodos para alcanzar y monitorizar el consenso cuando hay más de un revisor.

Independientemente del número de revisores, para asegurar que los criterios de evaluación que se propone utilizar pueden aplicarse consistentemente, es importante primero ponerlos a prueba mediante una prueba piloto sobre una muestra de artículos. Una muestra entre tres y seis artículos con diferentes niveles de riesgo de sesgo podría ser una muestra adecuada para tal fin.

¿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 pueden sesgar sus evaluaciones (Oxman 1993b), es posible que evalúen la validez de los estudios de manera más consistente que las personas sin experiencia sobre el tema (Jadad 1996). También pueden hacer valiosas apreciaciones distintas de las que podrían hacer personas que sólo tienen experiencia metodológica. Intuitivamente, sería aconsejable usar tanto expertos en el tema como no expertos y asegurar que todos tienen una comprensión adecuada de los aspectos metodológicos pertinentes.

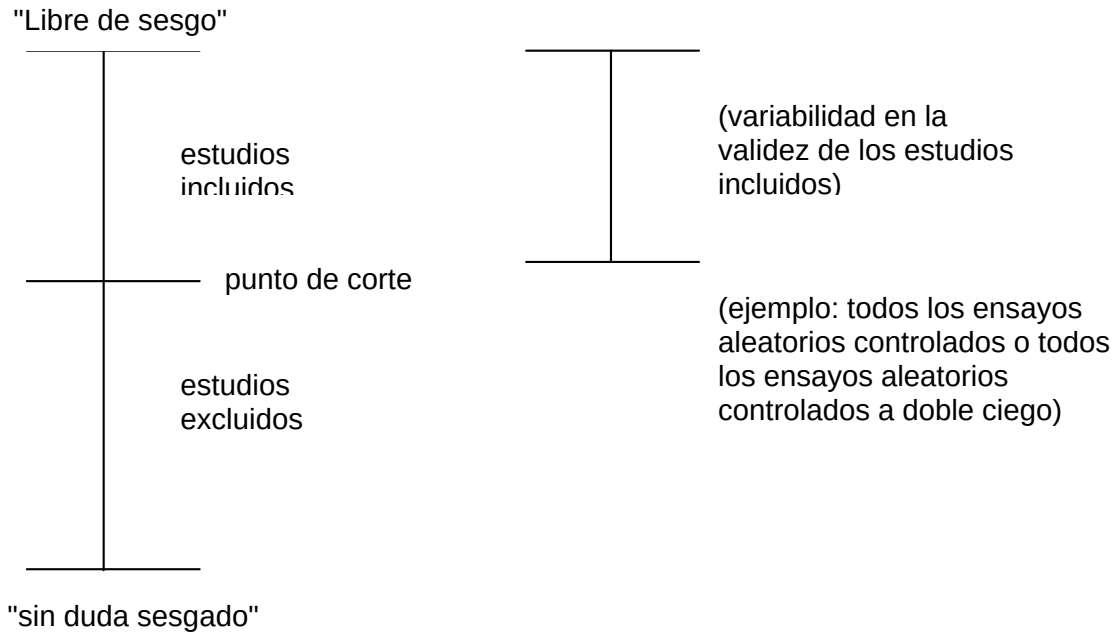
Los revisores también deben decidir si ocultan los nombres de autores, instituciones, publicaciones y resultados de un estudio a aquellas personas que evalúen la validez de los estudios, al momento de evaluar sus métodos. Algunos estudios empíricos indican que la evaluación a ciegas de los informes puede producir puntuaciones inferiores y más uniformes que las evaluaciones abiertas (Jadad 1996). Otra evidencia empírica sugiere que el cegamiento en las evaluaciones reporta pocos beneficios (Berlin 1997). No obstante, las evaluaciones con cegamiento requieren mucho tiempo. Los revisores deben sopesar sus beneficios potenciales frente a los costos ocasionados, cuando deciden si ciegan o no a los revisores. Actualmente se están llevando a cabo estudios para comparar evaluaciones abiertas y ciegas de la validez de los estudios y sus resultados podrán ayudar a guiar esta decisión.

6. 10. 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 de los estudios

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 estudios que previenen el sesgo de realización mediante el cegamiento de los participantes. La decisión sobre dónde situar el punto de corte para la inclusión se puede conceptualizar como la existencia de una línea continua que va desde los estudios “libres de sesgo” en un extremo, hasta los estudios “sin duda sesgados” en el orto, tal como se ilustra a continuación:



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 de inclusión. Con un punto de corte suficientemente elevado, cualquier variación en la validez de los estudios probablemente sea demasiado pequeña para que sea importante.

Hay diferentes métodos para explorar si la validez puede explicar las diferencias entre los resultados de los estudios incluidos (Detsky 1992). Se pueden utilizar gráficos visuales de los resultados ordenando los estudios según su grado de validez. Una segunda aproximación consiste en analizar subgrupos de estudios situados por encima del punto de corte metodológico, el cual preferiblemente debe ser especificado *a priori* en el protocolo de la revisión. Tanto si los resultados de los estudios son heterogéneos o no, este enfoque puede utilizarse mediante un análisis de sensibilidad a fin de determinar si los resultados globales coinciden con los que se obtienen cuando sólo se incluyen en el análisis los estudios con riesgo de sesgo bajo. Un tercer enfoque consiste en ordenar los estudios según su validez y combinar secuencialmente los resultados de cada estudio (un metaanálisis “acumulativo”), examinando el impacto en los resultados globales a medida que se van incluyendo ensayos cada vez de menor validez (ver sección 8.11.6).

Un cuarto enfoque consiste en utilizar métodos estadísticos para ponderar los estudios según su validez o utilizar una “metaregresión” para explorar la relación entre validez y magnitud del efecto entre estudios (ver sección 8.8.1). Los métodos estadísticos que combinan los resultados de los estudios suelen ponderar cada estudio por el inverso de la varianza de la medida de efecto estimada. En otras palabras, los estudios con resultados más precisos (intervalos de confianza más estrechos) tienen más peso.

También es posible ponderar los estudios según la validez, de manera que los estudios más válidos tiene más influencia en la medida resumen de efecto. La mayor

objeción a este enfoque es que no hay una base empírica para determinar qué peso debe asignarse a los distintos criterios de validez o para relacionar cuantitativamente las diferencias en las escalas de “calidad” disponibles con las diferencias en cuanto al riesgo de sesgo entre los estudios.

Es posible usar *RevMan* 4.0 para ordenar los estudios según el grado de ocultación o las valoraciones de validez “definidas por el usuario”. Se pueden realizar análisis de subgrupos sobre la base de las valoraciones de validez, aunque no se dispone de un test de significación estadística de las diferencias entre los subgrupos. *RevMan* no incluye una opción para ponderar los estudios según la validez metodológica y no es posible usar *RevMan 4.0* para hacer una metaregresión.

6. 11. Limitaciones de la evaluación de la calidad

Al evaluar la validez de los estudios surgen dos grandes dificultades. La primera es la publicación inadecuada de los ensayos clínicos (SORT 1994, Schulz 1994, WGRR 1994, Begg 1996). Se puede suponer que si algo no se informó, no se realizó. Sin embargo, no tiene por qué ser necesariamente así. Aunque es una tarea difícil, los revisores deben intentar obtener datos adicionales de los investigadores cuando sea necesario. La utilización de métodos estandarizados de publicación de los ensayos clínicos (SORT 1994, WGRR 1994, Begg 1996) puede facilitar la evaluación de la validez de los estudios en el futuro.

La segunda limitación, que en parte es una consecuencia de la primera, es la limitada evidencia empírica de la relación existente entre los parámetros propuestos para medir la validez y los resultados reales del estudio. Como se señalaba anteriormente, hay resultados empíricos que indican que, en general, tanto la ocultación inadecuada de las asignaciones como la ausencia de doble ciego producen sobreestimaciones de los efectos del tratamiento. Se necesita investigar más para establecer qué criterios utilizados en la evaluación de la validez son realmente factores determinantes de los resultados de los estudios. Una mejora en la presentación de los métodos puede facilitar tales investigaciones. Mientras tanto, los revisores deberían evitar el uso de las "puntuaciones de calidad" y la confianza indebida en las evaluaciones detalladas de calidad. Su uso no está sustentado por la evidencia empírica, puede ser lento y es potencialmente engañoso.

6.12 Referencias

Begg 1996. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637-9.

Berlin 1997. Berlin JA, on behalf of the University of Pennsylvania meta-analysis blinding study group. Does blinding of readers affect the results of meta-analyses? Lancet 1997; 350: 185-6.

Bero 1999. Bero L, Grilli R, Grimshaw J, Mowatt G, Oxman A, Zwarenstein M (editors). Effective Practice and Organisation of Care Module. In: The Cochrane Library, Issue 2, 1999. Oxford: Update Software.

CCSG 1978. The Canadian Cooperative Study Group. The Canadian trial of aspirin and sulfinpyrazone in threatened stroke. N Engl J Med 1978; 299:53-9.

Chalmers 1983. Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983; 309:1358-61.

Chalmers 1989. Chalmers I, Hetherington J, Elbourne,D, Keirse MJNC, Enkin M. Materials and methods used in synthesizing evidence to evaluate the effects of care during pregnancy and childbirth. In: Chalmers I, Enkin M, Keirse MJNC, editors. Effective Care in Pregnancy and Childbirth. Oxford: Oxford University Press, 1989; 39-65.

Colditz 1989. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. I: medical. Stat Med 1989; 8:441-54.

Detsky 1992. Detsky AS, Naylor CD, O'Rourke K, McGreer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. J Clin Epidemiol 1992; 45:255-65.

Emerson 1990. Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. Controlled Clin Trials 1990; 11:339-52.

Feinstein 1982. Feinstein AR, Horwitz RI. Double standards, scientific methods, and epidemiological research. N Engl J Med 1982; 307:1611-7.

Feinstein 1985. Feinstein AR. Clinical Epidemiology: The Architecture of Clinical Research. Philadelphia: Saunders, 1985: 39-52.

Gotzsche 1989. Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. Controlled Clin Trials 1989;10:3-56.

Horwitz 1979. Horwitz RI, Feinstein AR. Methodological standards and contradictory results in case-control research. Am J Med. 1979; 66:556-64.

Jadad 1996. 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.

Jüni 1999. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999; 282: 1054-60.

Karlowski 1975. Karlowski TR, Chalmers TC, Frenkel LD, Kapikian AZ, Lewis TL, Lynch JM. Ascorbic acid for the common cold: a prophylactic and therapeutic trial. JAMA 1975; 231: 1038-42.

Kleijnen 1997. Kleijnen J, Gotzsche P, Kunz RA, Oxman AD, Chalmers I. So what's so special about randomization? In: Chalmers, I, Maynard, A, editors. Non-Random Reflections on Health Services Research. London: BMJ, 1997; 93-106.

Kunz 1995. 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.

Kunz 1998. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ 1998; 317:1185-90.

Levine 1994. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, IV: how to use an article about harm. JAMA 1994; 271:1615-9.

Moher 1995. Moher D, Jadad A, Nichol G, Penman M, Tugwell T, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Controlled Clin Trials 1995; 16:62-73.

Moher 1996b. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: current issues and future directions. Int J Tech Assess in Health Care 1996; 12:195-208.

Moher 1998a. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609-13.

Oxman 1993b. Oxman AD, Guyatt GH. The science of reviewing research. Ann NY Acad Sci. 1993;703:125-33.

Reitman 1988. Reitman D, Chalmers TC, Nagalingam R, Sacks H. Can efficacy of blinding be documented by meta-analysis? Presented to the Society for Clinical Trials, San Diego, 23-26 May, 1988.

Sackett 1979b. Sackett DL. Bias in analytic research. J Chronic Dis 1979; 32:51-63.

Schulz 1994. Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. JAMA 1994;272:125-8

Schulz 1995. Schulz KF, Chalmers I, Hayes RJ, Altman D. Empirical evidence of bias. JAMA 1995; 273:408-12.

SORT 1994. The Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. JAMA 1994; 272:1926-31.

WGRR 1994. Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature. Ann Intern Med 1994; 121:894-5.

7. PROCESO DE EXTRACCIÓN DE DATOS

7. 1. Justificación de los formularios de recogida de datos

El formulario de recogida de datos es un puente entre lo que han informado los investigadores de los estudios primarios (como artículos de revistas, informes del proyecto, comunicaciones personales) y lo que, en último término, informa un revisor. El formulario de recogida de datos sirve al menos para tres funciones importantes. En primer lugar, el formulario de recogida de datos está directamente relacionado con la pregunta formulada en la revisión y con la evaluación que se ha planeado de los estudios incluidos, y por lo tanto, proporciona una representación visual de estos aspectos. Segundo, el formulario de recogida de datos es el registro histórico de multitud de decisiones (y cambios de decisiones) que ocurren en todo el proceso de revisión. Tercero: el formulario de recogida de datos es el almacén de datos del que surgirá el análisis.

Dado que los formularios de recogida de datos tienen funciones importantes, en su diseño hay que invertir tiempo y reflexión. Puesto que cada revisión es distinta, los formularios de recogida de datos varían. Sin embargo, hay una parte de la información que debe recogerse en todas las revisiones y los formularios pueden adaptarse de una revisión a otra.

7. 2. Formatos de extracción electrónicos *versus* formatos en papel

¿Los formularios de recogida de datos deben diseñarse en papel o en formato electrónico? Las hojas impresas pueden ser más fáciles de diseñar ya que los formatos electrónicos requieren conocimientos de programación. Por otro lado, la gran cantidad de datos de las revisiones que se incluyen en un gran número de estudios se almacena y recupera más fácilmente en formato electrónico que en papel.

Si se utilizan formularios electrónicos el proceso de extracción de la información de los estudios coincide con el de la entrada de datos. Este tipo de formularios también puede usarse para calcular variables o hacer conversiones sencillas (por ejemplo, libras y kilogramos) de datos que se presentan en diversos formatos en diferentes estudios. Tanto los formularios electrónicos como los de papel pueden diseñarse de manera que proporcionen un registro histórico de las decisiones y modificaciones que ocurran en el proceso de revisión.

Muchos revisores usan un proceso de extracción doble mediante el cual dos evaluaciones independientes de cada estudio pueden compararse y consensuarse si fuera necesario. Cuando se usan formularios impresos de recogida de datos, el proceso de comparación es sencillo: se usa uno de los formularios para marcar y corregir los errores y los desacuerdos. El proceso de comparar las extracciones dobles mediante formularios electrónicos es rápido, pero requiere el desarrollo y contraste de programas compatibles con la base de datos que se use.

Identificar y abordar los errores y los desacuerdos entre los revisores quizá sea más difícil con las versiones electrónicas que con las de papel. Esto se debe a la dificultad de comparar electrónicamente los campos o las áreas de los formularios que permiten respuestas flexibles. Algunos aspectos relacionados con la programación pueden dificultar más las modificaciones o los cambios en los formularios originales si se trabaja con formatos electrónicos. Un último inconveniente posible de los formularios

electrónicos es su compatibilidad con el programa *Review Manager* (*RevMan*) que se usa para generar y almacenar la revisión final. Aunque hay distintas formas de transferir los datos de formularios electrónicos a RevMan, esto puede no ser tan simple e idealmente debe ser planeado con anterioridad.

Si se usa un formato electrónico, deben tenerse en cuenta las siguientes indicaciones. En primer lugar, no debe programarse el formulario electrónico hasta que se ha diseñado, experimentado y depurado una copia del mismo en papel. Idealmente, la prueba piloto debería incluir más de un revisor y varios artículos. En segundo lugar, cuando se diseña el formulario de recogida de datos, hay que considerar las preferencias de la persona que entra los datos, hay que estructurarlo de manera lógica y hay que codificar las respuestas de la manera más consistente y sencilla posible. En tercer lugar, cuando se elige una base de datos electrónica o una hoja de cálculo, hay que verificar si se puede crear un archivo electrónico compatible con RevMan. En cuarto y último lugar, no hay que olvidarse de crear mecanismos de control de calidad de la entrada de datos y de corrección cuando se hayan producido errores.

7. 3. Manejo de datos y *software*

Existe una amplia variedad de *software* y programas de manejo de datos para ayudar en el proceso de revisión sistemática. Para los formularios electrónicos de recogida electrónica de datos se puede utilizar *software* tipo hoja de cálculo, como QuatroPro, Excel y Lotus o programas de bases de datos como FoxPro o DataEase. Si se han planeado análisis no disponibles en RevMan, puede utilizarse s*oftware* como DBMSCOPY para transformar los archivos de bases de datos en archivos compatibles con otros programas (ver sección 8. 8).

7. 4. Componentes clave de un formulario de recogida de datos

Hay más de una forma correcta de diseñar un formulario de recogida de datos y se realizan algunas sugerencias basadas en la experiencia. Al adaptar o diseñar un formulario de recogida de datos, los revisores deben considerar en primer lugar la cantidad de información que desean recoger. Una recogida de información excesivamente detallada puede dar lugar a formularios más largos que los informes del estudio original, que sean tediosos y aburridos de completar y que requieran un tiempo excesivo para los revisores. Por otro lado, si los formularios no se detallan suficientemente y se omiten datos clave, los revisores pueden tener que resumir de nuevo los estudios usando formularios suplementarios. Tener que revisar de nuevo un estudio es lento y frustrante.

7. 4. 1. Información acerca de las referencias de estudios y revisores

Dado que los formularios de recogida de datos se pueden adaptar de una revisión a otra y que algunos revisores participan en múltiples revisiones, es necesario registrar el título de la revisión y el nombre del revisor que está resumiendo los datos. Es útil dejar un espacio después del título para que los revisores puedan escribir notas específicas sobre el estudio objeto de resumen. Esto evita colocar notas, preguntas o recordatorios en la última página, donde tienen menos probabilidad de ser vistos. Las notas importantes pueden introducirse en el programa RevMan en la columna "notas" de la tabla “Características de Estudios Incluidos”, o en el texto mismo de la revisión.

A cada revisión Cochrane se le asigna un identificador único. Este identificador debe estar situado a continuación del título en el formulario de recogida de datos. Ocasionalmente, tienen que revisarse los formularios. Si en el formulario se codifica la fecha de revisión o el número de la versión se reducen las posibilidades de usar por error un formulario antiguo.

A cada uno de los estudios incluidos debe asignársele un identificador que se utiliza en el programa *RevMan*. Es posible que los revisores deban recoger datos de múltiples informes del mismo ensayo. 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 ser codificada de la misma manera que la información publicada.

7. 4. 2. Verificación de la elegibilidad de un estudio

Aunque la búsqueda y el proceso de selección deberían haber descartado la mayoría de estudios no elegibles, es bueno comprobar la elegibilidad de los estudios en el momento del resumen o de la recogida de datos. La verificación de la elegibilidad del estudio debe hacerse pronto ya que el resto del formulario se refiere a los estudios que reúnen los criterios de inclusión, y extraer datos de estudios que después se excluirán significa malgastar recursos.

Las revisiones Cochrane incluyen una tabla de estudios excluidos para aquellos estudios que parecen reunir los criterios de inclusión pero que, tras un examen más minucioso, fueron excluidos.

En el formulario de recogida de datos puede incluirse un apartado para codificar las razones por las que se excluyeron tales estudios. Por ejemplo, supongamos que un revisor sólo quiere incluir en la revisión los ensayos completamente aleatorizados. Una pregunta de verificación en el formulario de recogida de datos 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 en *RevMan* como motivo de exclusión.

7. 5. Características de los estudios

Para evaluar cada estudio hay que codificar algunas de sus características específicas. Estas pueden agruparse tal como se estructura la información que se entrará en el programa RevMan: métodos, participantes, intervenciones y medidas de resultado. La información sobre los participantes puede incluir detalles relacionados con el ámbito de estudio y los criterios diagnósticos para la condición de interés.

El desarrollo de esta parte del formulario de recogida de datos requiere reflexión y hacer una prueba piloto. Los datos que se recogen deberían estar vinculados directamente al tema o temas de la revisión y a las estrategias de análisis planificadas. Los datos deberían recogerse en un formato que facilite la entrada en el programa RevMan.

7. 5. 1. Métodos

Los diferentes métodos de investigación pueden influir en las medidas de resultado de los estudios, introduciendo sesgos en los resultados. Por ejemplo, según se discutió

en la sección 6, es importante que la asignación a los tratamientos se oculte adecuadamente. Cuando se introduzca en RevMan la información acerca de los estudios individuales, será necesario codificar la ocultación de asignaciones como “adecuada” (A), “poco clara” (B), “inadecuada” (C) o “no se usó” (D). Los formularios de recogida de datos deben reflejar estas valoraciones.

Otras características de los métodos que pueden ser relevantes incluyen 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 co-intervenciones y otros factores potenciales de confusión. El apartado de métodos del formulario de recogida de datos debe incluir los criterios de validez utilizados.

7. 5. 2. Participantes

Las características de los participantes pueden variar sustancialmente entre los estudios y algunos Grupos Colaboradores de Revisión (GCR) han desarrollado normas sobre qué características deben recogerse. Los elementos que deben recogerse son aquellos que pueden afectar a los resultados de los estudios o ayudar a los usuarios que evalúan la aplicabilidad de los resultados. Por ejemplo, si el revisor tiene motivos para creer que hay diferencias importantes en el efecto del tratamiento entre diversos grupos étnicos, esta información debe recogerse.

Si los efectos de los tratamientos se consideran constantes en los distintos grupos étnicos, y si además, la información sobre el grupo étnico no ayudaría a aplicar los resultados, no debe ser notificada. Algunas características que son útiles para evaluar la aplicabilidad incluyen la edad y el sexo. Otras variables sociodemográficas como el nivel de educación o las variables que tratan la presencia de comorbilidades también son importantes.

Los ámbitos de los estudios deben evaluarse si se piensa que pueden influir en los efectos del tratamiento o en su aplicabilidad. Los ámbitos que aparecen con más frecuencia en los estudios de intervención sanitaria son: hospitales de agudos, departamentos de urgencias; consultas o consultorios; establecimientos de larga estancia, como residencias y casas de convalecencia; o, comunidades geográficas.

A veces los estudios se realizan en regiones geográficas distintas con diferencias culturales importantes que podrían afectar a la prestación de una intervención y sus resultados. A veces las diferencias temporales indican diferencias tecnológicas importantes. Si tales elementos son importantes para la interpretación de la revisión, deberían evaluarse.

Los criterios diagnósticos utilizados para definir el problema de salud de interés pueden ser una fuente importante de variabilidad clínica y deberían describirse. Por ejemplo, en una revisión sobre terapia farmacológica para la insuficiencia cardíaca congestiva, es importante conocer cómo se estableció en cada estudio la definición y gravedad de la insuficiencia cardíaca (es decir, insuficiencia sistólica, disfunción diastólica, disfunción sistólica grave con fracciones de eyección de < 20%, etc.). De manera similar, en una revisión de tratamiento antihipertensivo, es importante describir los niveles básicos de presión arterial de los participantes.

7. 5. 3. Intervenciones

Debe describirse la intervención y de qué forma se realizó. Para los ensayos con tratamientos farmacológicos, deberían evaluarse las vías de administración (oral, intravenosa, etc.), las dosis y el tiempo (durante las 24h a partir del diagnóstico, por ejemplo). La duración del tratamiento, si fue diferente al periodo de seguimiento y no se registró en el apartado de métodos, también puede incluirse aquí. En las intervenciones complejas como los estudios que evalúan las psicoterapias, los enfoques educacionales 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.

Para los ensayos que no utilizan placebos y los que evalúan intervenciones complejas, también es importante registrar la información referente a lo que se dio al grupo control. Más adelante, esta información ayudará a tomar decisiones acerca de si es razonable combinar los datos entre estudios, dado que una heterogeneidad considerable en cuanto a lo que reciben los grupos control puede ser una razón para no combinar los estudios, o para hacer un análisis de sensibilidad.

7. 5. 4. Medidas de resultado y resultados de los estudios

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

El revisor necesita combinar la información necesaria sobre medidas de resultado para responder a la pregunta o preguntas de la revisión, a partir de los informes de los estudios. Para evitar errores, las medidas de resultado deberían recogerse en el formato que se presentó en el estudio original y transformarse en un paso posterior. En los ensayos cruzados y los ensayos con medidas de resultado obtenidas en diversos períodos de seguimiento, es necesario decidir qué resultados se van a evaluar (ver secciones 8.11. 5 y 8.9. 1 respectivamente).

Los revisores deben valorar la posibilidad de diseñar los formularios para que se ajusten a las tablas de datos del RevMan. Por ejemplo, si el revisor quiere usar datos continuos, se necesita la siguiente información para cada grupo de comparación: el número de participantes, la media y la desviación estándar. Sin embargo, estos valores pueden ser insuficientes porque hay gran variabilidad en la información que proporcionan los investigadores

Por ejemplo, los investigadores pueden presentar un intervalo de confianza para la diferencia de medias y no presentar la desviación estándar, o pueden presentar sólo el valor de una prueba estadística (t, F, ji cuadrado, etc.) o el valor p. Los formularios de recogida de datos deben ser flexibles para tratar esta variabilidad en la valoración de las medidas de resultado. Para más detalles relacionados con qué información de medidas de resultado es necesaria según el tipo específico de análisis, véase la sección 8.

7. 6. Formato de la codificación e instrucciones para los codificadores

Es sumamente importante utilizar un sistema de codificación correcto. La codificación no debe ser complicada 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 recogida de datos. Hay disparidad en las preferencias de colocación de las instrucciones. Una propuesta es insertar la instrucción a continuación o cerca de la variable que se codificará. En algunos casos, las instrucciones pueden ser extensas y quizá deban colocarse en otra página. Independientemente del método usado (lo más probable es que sea una mezcla), es muy importante que los revisores se entrenen en el uso del formulario y reciban o proporcionen algún tipo de formación si el formulario fue diseñado por otra persona diferente a aquella que lo va a utilizar.

7. 7. Prueba piloto y revisión de formularios

Todos los formularios deberían probarse en un estudio piloto con una muestra representativa de los estudios que se revisan. Esta prueba puede identificar las variables que no se necesitan o a las que faltan datos. Las personas que resumen los datos pueden sugerir cambios en aquellas instrucciones de codificación confusas o incompletas (por ejemplo, pueden no estar descritos todos los tipos de respuestas). Cuando participan múltiples revisores en un proyecto, debería haber consenso entre ellos antes de modificar el formulario para evitar malentendidos o desacuerdos posteriores. Según la complejidad de la revisión y la experiencia de la prueba piloto, pueden ser necesarias pruebas piloto adicionales.

A veces pueden surgir problemas con el formulario de recogida de datos una vez se ha completado la prueba piloto y se ha revisado el formulario. De hecho, es poco frecuente que un formulario de recogida de datos no requiera modificaciones después de la prueba piloto. Cuando tienen que hacerse cambios en el formulario o en las instrucciones de codificación, se debe asegurar la corrección de los formularios de aquellos estudios que ya se han revisado. En algunos casos, quizá sólo sea necesario esclarecer las instrucciones de codificación sin modificar el formulario de recogida de datos.

7. 8. Fiabilidad de la recogida de datos

La fiabilidad se refiere al grado en que diferentes personas coinciden en la revisión de un estudio. Por ejemplo, ¿cada revisor estuvo de acuerdo sobre la presencia de comorbilidad entre los participantes de un ensayo específico? ¿Estuvieron de acuerdo sobre los datos de las medidas de resultado en cada grupo de comparación? Cuando más de una persona examina los datos, inevitablemente habrá desacuerdos. Los grupos de múltiples revisores deben planificar cómo compararán la información de sus formularios de recogida de datos y cómo alcanzarán el consenso en caso de desacuerdo.

El consenso puede lograrse mediante discusión entre revisores o utilizando un árbitro adicional independiente. También es importante planificar cómo se registrará el acuerdo de "consenso". Como mínimo hay tres posibilidades:

1) Usar el formulario de uno de los revisores y registrar los cambios en tinta roja después del consenso.
2) Usar otro formulario impreso.
3) Introducir sólo los datos de consenso en formato electrónico.

Para evaluar la fiabilidad de la codificación es esencial mantener separada la información sobre el consenso.

Quizá no sea importante analizar formalmente la confiabilidad de todos los datos recogidos. Por ejemplo, un revisor puede optar por restringir la evaluación de la fiabilidad a las medidas de resultado y las evaluaciones de validez de los estudios. No hay un estándar fijo sobre el nivel de fiabilidad adecuado o cómo evaluar la fiabilidad. Sin embargo, es importante analizar la fiabilidad en todo el proceso de recogida de datos. Por ejemplo, si después de alcanzar el consenso en los primeros estudios, los revisores observan discrepancias frecuentes en datos específicos, probablemente es necesario modificar las instrucciones de codificación. Los revisores pueden mostrar cambios en la codificación (*coder drift* o cambio temporal en la forma de codificar la información), indicando posiblemente la necesidad de reentrenamiento o de recodificación. Esto sólo puede identificarse cuando la fiabilidad se analiza en todo el proyecto.

7.9 EXTRACCIÓN CIEGA DE LOS DATOS

BAJO CONSTRUCCION. Se está preparando una sección acerca de si la obtención de los datos debe ser ciega; por ejemplo, con respecto a los autores, a las revistas y a los resultados al momento de evaluar la calidad. Aunque existe alguna evidencia que indica que las evaluaciones ciegas de la calidad de los ensayos pueden ser más confiables y diferentes de aquellas evaluaciones que no son ciegas (Jadad 1996, Moher 1998b), el cegamiento es difícil de lograr, es dispendioso y puede que no altere sustancialmente los resultados de una revisión (Berlin 1997a, Berlin 1997b).

7.10 OBTENCIÓN DE DATOS DE LOS INVESTIGADORES

A menudo, los revisores consideran que no son capaces de obtener toda la información de interés a partir de los informes publicados, ya sea referente a los detalles del estudio o bien con relación a sus resultados numéricos. En tales circunstancias, los revisores necesitan definir cómo van a obtener la información que hace falta. Pueden querer contactar a los investigadores originales y por ejemplo, deben considerar si van a contactarlos con una solicitud abierta, con el envío de un formato estandarizado de obtención de datos, con una solicitud de datos individuales de pacientes (ver sección 11), o si van a hacer alguna otra cosa.

7.1. Referencias

Jadad 1996. 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.

Moher 1998b. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses?. Lancet 1998; 352:609–13.

Berlin 1997a. Berlin JA, Miles CG, Crigliano MD. Does blinding of readers affect the results of meta-analyses? Results of a randomized trial. Online J Curr Clin Trials 1997.

Berlin 1997b. Berlin JA. Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. Lancet 1997; 350: 185-6.

8. ANÁLISIS Y PRESENTACIÓN DE RESULTADOS

8.1. ¡No comience por aquí!

Por favor lea las secciones 3 a 7 antes de leer esta sección. A veces, al iniciar una revisión uno puede sentirse tentado de pasar directamente al análisis. Esto es arriesgado si no se ha prestado la atención suficiente a la formulación del problema (sección 3 y sección 4), a la identificación, selección y evaluación crítica de los estudios (secciones 5 y 6) y a la recogida de datos (sección 7).

8. 2. Planificación del análisis

La razón para realizar las revisiones de forma sistemática es asegurar la validez de los resultados. El papel del análisis estadístico en las revisiones puede estar menos claro. A veces, el uso de la estadística (metanálisis) puede parecer - y quizá sea - más un obstáculo que una ayuda para los que están familiarizados con estas técnicas. Existe también el riesgo de prestar demasiada atención a la línea estadística inferior o al rombo o diamante situado al final del gráfico que resume los resultados de estudios individuales en un metanálisis. A veces, los revisores y los usuarios de las revisiones, con las prisas de llegar al rombo o diamante, descuidan la importancia de la reflexión y de las consideraciones hechas en los pasos anteriores al análisis, en el análisis mismo y en el apartado de conclusiones. En esta sección se tratará de lo que supone el análisis de una revisión, el papel de la estadística en este proceso y la necesidad de tener buen criterio por parte del revisor y del usuario de la revisión.

8. 2. 1. ¿Por qué realizar un metanálisis en una revisión?

Para preparar una revisión, se deben obtener datos de los estudios individuales y se pueden usar métodos estadísticos para analizar y resumir estos datos. Si se usan apropiadamente, tales métodos proporcionan una herramienta potente para llegar a unas conclusiones que tengan sentido y ayudan a evitar errores de interpretación.

Cuando en una revisión no se usan métodos estadísticos, es fácil cometer el error de comparar el número de estudios "positivos" con el número de estudios "negativos". Este "recuento de votos" debe evitarse. Es poco confiable ya que un estudio puede considerarse "positivo" o "negativo" dependiendo de cómo interpreten los resultados los revisores y porque no tiene en cuenta el peso relativo de la evidencia fiable proporcionada por cada estudio. Al contar los votos también se tiende a pasar por alto los efectos pequeños pero clínicamente importantes, en particular al contar como "negativos" los estudios con resultados que no son estadísticamente significativos (Cooper 1980, Antman 1992).

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 una revisión sean válidos, como ocurre también en un estudio primario. Como cualquier herramienta, los métodos estadísticos pueden utilizarse mal.

8. 2. 2. ¿Qué implica un metanálisis?

Aunque el uso de métodos estadísticos en las revisiones puede ser sumamente útil, el elemento más esencial de un análisis es el exámen detenido de 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. Posteriormente, se pueden calcular de una manera sistemática, los estimadores del efecto a lo largo de los estudios para concluir cuánta confianza debe depositarse en estos estimadores y para investigar las diferencias entre los estudios.

8. 2. 4. ¿Qué comparaciones deben hacerse?

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 o las hipótesis que se plantean cuando se formula el objetivo de la revisión (ver sección 4). 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, y que variaciones importantes en la intervención sólo se puedan observar después de recoger los datos.

Las decisiones acerca de qué resultados de los estudios son suficientemente similares para poder agruparlos requieren un buen entendimiento del problema que trata la revisión y la elaboración de juicios por parte del revisor y del usuario de la revisión. En la sección 4 se tratan las consideraciones acerca del planteamiento de las preguntas que la revisión aborda. Esencialmente, se pueden aplicar las mismas consideraciones para decidir qué comparaciones se deben hacer, qué resultados se deben combinar y qué características clave hay que tener en cuenta (del diseño de los estudios, de los participantes, las intervenciones y los resultados) al investigar la heterogeneidad (variabilidad en los efectos).

Estos aspectos deben tratarse cuando se elaboran comparaciones en RevMan y cuando se decide qué información se va a colocar en la Tabla de Características de Estudios Incluidos.

8. 2. 5. ¿Qué resultados de los estudios 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 el caso de variables dicotómicas (resultados que están presentes o ausentes, como la mortalidad), los datos que deben resumirse son el número de personas que experimentaron el suceso o resultado en cada grupo de comparación, y el número total de personas en cada grupo. Debe considerarse el orden en que se presentarán

los resultados (por año, alfabéticamente según el identificador del estudio, según los resultados). El orden puede modificarse tanto en el RevMan como en la *Cochrane Database of Systematic Reviews* (*CDSR*), pero por defecto, la presentación de cada resultado en cada revisión es determinada por los revisores.

Dado que la ocultación de la asignación es un criterio de validez que afecta a todos los resultados dentro de un estudio y que hay evidencia para apoyar su uso, este criterio se asigna cuando se introducen los estudios en RevMan y constituye una característica por la cual se pueden ordenar los estudios en los análisis. Los revisores también pueden ordenar los estudios de acuerdo con otras evaluaciones de la validez u otras características definidas por el usuario.

En el caso de variables continuas (como tensión arterial) los datos que deben resumirse son el número de personas en cada grupo y la media y la desviación estándar de la medida de resultado en cada grupo. Lamentablemente, a menudo las desviaciones estándar no se presentan. Aunque las desviaciones estándar pueden obtenerse a partir de otras medidas de variación informadas (por ejemplo, los intervalos de confianza), esto no siempre es posible. Las opciones para manejar este caso y otros tipos de datos que faltan (*missing*) se tratan más adelante, en la sección 8. 11. 4.

A veces los resultados de una variable continua en algunos estudios se presentan como cambios (media de las diferencias entre mediciones pre y post-intervención), y en otros estudios como resultados post-intervención. Cuando se da esta situación no hay una única solución óptima para el resumen de los resultados. Como regla general, en primer lugar los revisores deberían pensar qué manera de presentar los resultados es más útil y fácil de comprender para los usuarios de sus revisiones. Si es posible, deben transformarse los resultados y deben buscarse los datos de los investigadores para que todos los resultados del estudio puedan presentarse de la misma forma.

Hay diversas opciones disponibles para imputar resultados, pero todas ellas dependen de ciertas hipótesis, por lo cual deben usarse con precaución y únicamente si hay indicios de que las hipótesis se cumplen. Deberían hacerse análisis de sensibilidad para observar las consecuencias de cambiar las hipótesis (ver sección 8. 10).

En algunas ocasiones, puede suceder que los revisores encuentren los datos de una misma medida de resultado en forma dicotómica en algunos estudios y continua en otros. Por ejemplo: los resultados en las escalas de depresión pueden presentarse como una media o como el porcentaje de pacientes que estaban deprimidos en cierto periodo después de una intervención (una puntuación superior a un punto de corte determinado). A menudo, este tipo de información es más fácil de comprender y es más útil cuando se expresa en forma dicotómica.

Sin embargo, la decisión de establecer un punto de corte puede ser arbitraria y siempre se pierde información cuando las variables continuas se transforman en dicotómicas. Hay varias opciones para tratar de combinar datos dicotómicos y continuos. En general, es útil resumir los resultados de todos los estudios válidos y relevantes de manera similar, pero esto no siempre es posible.

Puede ser posible obtener los datos que faltan de los investigadores. Si no, puede ser útil resumir los datos en tres grupos: colocar los datos continuos en una Tabla de Datos Continuos, los datos dicotómicos en una de Tabla de Datos Dicotómicos y otros tipos de datos en una Tabla para Otros Datos. Otra opción es presentar todos los datos como media estandarizada de las diferencias. Por otra parte, el *log-odds ratio*

puede usarse como una medida de la diferencia entre los tratamientos (Whitehead 1999).

Pueden encontrarse otras situaciones donde no es adecuado resumir los datos en una tabla de datos continuos o dicotómicos. Un ejemplo sería cuando estudios que abordan el mismo tema, utilizan medidas de resultado distintas. Por ejemplo, si uno está interesado en sintetizar la investigación sobre los efectos de la auditoría y la retroalimentación sobre la práctica profesional, puede ser útil resumir mediante medidas objetivas de desempeño todos los estudios válidos que comparan los efectos de la auditoría y la retroalimentación contra ninguna intervención, aunque los tipos de profesionales, los entornos y los resultados varíen sustancialmente de un estudio a otro.

Aunque en este caso puede ser posible transformar los resultados en diferencias estandarizadas de promedios, estas medidas pueden ser difíciles de interpretar. Quizá también es posible estandarizar los resultados de otra manera (por ejemplo, un cambio porcentual). Aún así, seguramente las personas que toman decisiones prefieren un resumen de los resultados en "unidades naturales" que sean intuitivas.

Si se han obtenido los datos individuales de los pacientes incluidos en los estudios de la revisión, esto puede ser analizado fuera de *RevMan* para obtener estadísticas de resumen para cada estudio. Estas medidas de resumen, junto con el número de eventos y de participantes en cada grupo se introducen en una Tabla de Datos de pacientes individuales en *RevMan.* Después, el programa de análisis utiliza las estadísticas de resumen que se introdujeron y no los números crudos.

No hay un conjunto de datos estándar que deba incluirse en la Tabla de Datos para Otros tipos de datos; es decir datos que no encajan claramente en las Tablas de Datos Dicotómicos, Continuos o Datos Individuales de Pacientes. En general, hay cinco tipos de información útiles para resumir los resultados y facilitar su análisis:

- Variables clave que caracterizan a los participantes.
- Variables clave que caracterizan las intervenciones.
- Variables clave que caracterizan las medidas de resultado.
- Resultados en unidades naturales que se comprendan fácilmente.
- Resultados estandarizados, si pueden calcularse de forma que faciliten las comparaciones entre estudios sin que creen confusión.

Los tres primeros tipos de información son "variables independientes" que pueden ayudar a explicar los resultados. Un cuarto tipo de variable independiente que podría incluirse también es la validez interna de cada estudio. La elección de variables específicas depende de la comprensión del problema que trata la revisión. Sin embargo estas variables deberían corresponder a las hipótesis *a priori* que se hayan incluido en el protocolo sobre diferencias importantes entre los estudios.

Si los estudios identificados no contribuyen con datos para un resultado particular, el revisor debe considerar si esta ausencia de datos podría o no estar sesgada. Por ejemplo, algunos resultados no se habrían recogido en ciertos estudios porque no había medios relevantes para hacer mediciones en el momento en que se hizo el estudio, de manera que la razón que explica su ausencia no puede estar influenciada por los resultados del estudio. Si embargo, si el estudio recogió datos relevantes sobre resultados pero no se encuentran disponibles para la revisión, esto puede ser debido a sesgos. El revisor debe decidir si debe incluir aquellos estudios que no tienen datos junto con la comparación asociada en la revisión. La inclusión de dichos estudios ayuda a informar al lector de cuánta evidencia se pierde para el resultado de

interés. Esto es particularmente importante si los datos relevantes se recogieron en un estudio, pero no se encuentran disponibles.

8. 3. Resumen de los efectos de los estudios

Un objetivo de la mayoría de las revisiones Cochrane es proporcionar una estimación fiable de los efectos de una intervención, basada en una media ponderada de los resultados de todos los estudios relevantes.

Generalmente, a cada estudio se le asigna como peso el inverso de su varianza; es decir, a las estimaciones más precisas (de estudios más amplios y con más sucesos), se les da más peso (Laird 1990). También es posible dar más o menos peso a los estudios con base a otros factores como su calidad metodológica, pero esto se hace raramente (Detsky 1992).

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 los demás participantes en el mismo estudio.

A veces es posible encontrar metanálisis en los que se calcula un "tamaño del efecto" para cada grupo de intervención en cada ensayo aleatorio (con base en la diferencia de resultados antes y después de la intervención), y los grupos de diferentes estudios se comparan directamente entre sí. De esta manera, cada grupo de intervención se vuelve el equivalente de un estudio de antes y después y se pierden los beneficios de la aleatorización original en cada ensayo. Los resultados de cualquier metanálisis de este tipo necesitan interpretarse con mucho cuidado y no debe considerarse como un metanálisis de ensayos aleatorios. El programa RevMan no incluye esta aproximación metodológica para la síntesis de los datos y debería evitarse.

8. 4. Métodos estadísticos disponibles en el programa RevMan y en la CDSR

Los metanálisis utilizan diversas técnicas. Lamentablemente, no hay una única técnica "correcta". La elección de la técnica depende del tipo de datos analizados. Afortunadamente, como para otros usos de la estadística, para los revisores es más importante una comprensión conceptual de los principios que el conocimiento detallado de las técnicas específicas. Los métodos estadísticos disponibles en RevMan, se mencionan a continuación:

Tipo de datos	Resumen estadístico	Modelo	Método
Dicotómicos	*odds ratio* (O-E)	efectos fijos	Peto Mantel-Haenszel
		efectos aleatorios	DerSimonian y Laird
	Riesgo relativo	efectos fijos efectos aleatorios	Mantel-Haenszel DerSimonian y Laird

	Diferencia de riesgo	efectos fijos efectos aleatorios	Mantel-Haenszel DerSimonian y Laird
Continuos	Diferencia de medias ponderada Diferencia estandarizada de medias	efectos fijos efectos aleatorios efectos fijos efectos aleatorios	inverso varianza DerSimonian y Laird inverso varianza DerSimonian y Laird
Datos de pacientes individuales	*Odds ratio* (O-E)	efectos fijos	Peto

8.4.1. Datos dicotómicos

Para datos dicotómicos (o binarios) hay varios pros y contras en cada uno de los estadísticos de resumen disponibles: *odds ratio*, riesgo relativo (o efecto relativo), y diferencia de riesgo (ver apéndice 8a). Aunque en metanálisis el *odds ratio* se ha usado de manera generalizada, existen preocupaciones relacionadas con la posibilidad de que sea interpretado incorrectamente (Sinclair 1994). El *odds ratio* tiene ventajas estadísticas en relación a su distribución muestral y su idoneidad en los modelos (Fleiss 1993), pero estas ventajas no siempre son importantes en metanálisis. Es más, no hay ninguna base sólida para suponer que el *odds ratio* (OR) varíe menos entre estudios que otros estadísticos de resumen (Sinclair 1994). El *odds ratio* y el riesgo relativo son similares si el resultado es relativamente raro.

El método estadístico utilizado y la presentación de los resultados pueden considerarse temas independientes. Por ejemplo, los *odds* ratio y los riesgos relativos son medidas relativas y pueden utilizarse para combinar estudios, pero la diferencia de riesgos es una medida absoluta que es mejor utilizar cuando se aplican los resultados a una situación sanitaria específica (Egger 1997a). El "número necesario a tratar" (NNT) (Laupacis 1988), es habitualmente una manera útil de expresar de nuevo los resultados. Puede obtenerse fácilmente a partir de la diferencia de riesgos y de manera más difícil a partir del riesgo relativo o del *odds ratio.* Nunca deben realizarse metanálisis que usen NNT de estudios individuales.

Las medidas absolutas pueden ser más informativas que las medidas relativas porque reflejan el riesgo basal y el cambio de riesgo con la intervención. Cuando se presentan medidas absolutas - como el NNT - es útil dar la información para un intervalo de riesgos basales (por ejemplo, para el intervalo observado en los estudios incluidos) (Smeeth 1999). Al mismo tiempo, la obtención de intervalos de confianza para medidas absolutas es un asunto problemático (Altman 1998), en parte porque también varían con el riesgo en la línea de base. Igualmente, los NNTs son específicos de un periodo de seguimiento, y solamente deben calcularse cuando se combinan ensayos que tienen periodos de seguimiento similares (Smeeth 1999). Esta información debe incluirse en la revisión. RevMan 4.0 no dispone de un comando para calcular los NNT, pero pueden encontrarse métodos para calcular los NNT en el Apéndice 8b.

A menudo es deseable describir los resultados con una medida relativa (*odds ratio*, riesgo relativo o reducción del riesgo relativo) y con una medida absoluta (NNT o diferencia de riesgos). En general, debe utilizarse en lo posible el riesgo relativo.

8.4.2. Datos continuos

Para datos continuos, la diferencia de medias ponderada debe usarse cuando los resultados se miden de manera estándar en los diferentes estudios. Esta medida tiene la ventaja de resumir los resultados en unidades naturales que se entienden

fácilmente. En algunas ocasiones, puede ser conveniente resumir los resultados de los estudios con indicadores de resultado conceptualmente iguales pero medidos de maneras distintas (por ejemplo, escalas diferentes de medición en esquizofrenia). En estas circunstancias pueden usarse las diferencias de medias estandarizadas. Dado que este enfoque permite combinar resultados de estudios sumamente distintos, los revisores deben ser especialmente cautelosos al decidir si este análisis tiene sentido y al interpretar los resultados.

Los revisores deben ser conscientes de la posibilidad de que haya datos asimétricos y de los problemas potenciales que conllevan. Para detectar una asimetría en aquellas mediciones que siempre son positivas (como tensión arterial), divida la media por la desviación estándar. Si el resultado es menor que 1.64, hay algún tipo de asimetría positiva.

8.4.3. Datos individuales de pacientes

Para los datos individuales de pacientes, las estadísticas de resumen se calculan para cada estudio mediante un nuevo cálculo de los datos crudos. Este análisis no es posible hacerlo en RevMan, y debe realizarse en un programa de análisis independiente con la ayuda de un estadístico. Después pueden introducirse en RevMan las estadísticas de resumen y combinarlas mediante el *Odds Ratio* de Peto.

8.4.4. Análisis de efectos fijos *versus* de efectos aleatorios

Existen diferentes técnicas para resumir los resultados de los estudios, y, lamentablemente hay desacuerdo en relación a cuál de ellas es mejor o acerca de si los resultados deben combinarse del todo en algunas situaciones, como cuando existe una heterogeneidad importante entre los resultados de los estudios (Schulz 1993). (La heterogeneidad se discute en la sección 8.6). El principal desacuerdo que surge en los metanálisis es si se incorpora la variación entre los estudios (heterogeneidad) al calcular las medidas combinadas de la magnitud del efecto. Si la variación entre los estudios es baja (es decir, la prueba de heterogeneidad arroja un valor P grande, mayor a 0.10), la decisión producirá pocas diferencias en los resultados. Sin embargo, si la variación entre los estudios es significativa, un análisis que ignore esta variación (un modelo de “efectos fijos”) proporcionará un intervalo de confianza más estrecho que uno que no lo hace (un modelo de “efectos aleatorios”). Para cada modelo, existen situaciones en las cuales el resultado sea contrario a lo que se intuye, de manera que puede ser apropiado analizar los resultados con ambos métodos. Fuera de esto, existen algunas bases comunes por lo menos en cuatro puntos.

En primer lugar y de manera más importante, los revisores deben alertar a los lectores de su revisión cuando hay una heterogeneidad importante, y deben invitarlos a que sean cautelosos en la interpretación de los resultados combinados. Inclusive podrían decidir que la heterogeneidad es de tal magnitud, que sería preferible no combinar los resultados de los estudios individuales. Los revisores deben intentar explorar las causas de la heterogeneidad y tratar de explicarla, pero cualquier interpretación debe ser cautelosa puesto que estos análisis se hacen habitualmente *a posteriori*. Idealmente, deben especificarse *a priori* dentro del protocolo de la revisión, los factores que podrían conducir a diferencias entre los resultados de estudios individuales. Estos pueden incluir diferencias en la definición de las intervenciones, los participantes, las maneras de evaluar los resultados o la calidad metodológica de los estudios. Las diferencias en la calidad metodológica podrían producir sesgos y acompañar a la heterogeneidad. Si existen diferencias importantes en la calidad de

los estudios, los revisores deben considerar el efecto que tiene excluir de la revisión aquellos estudios de peor calidad. Los análisis de subgrupos, los cuales pueden emplearse para examinar las diferencias entre los estudios, se discuten en la sección 8.7. (Yusuf 1991, Oxman 1992)

En segundo lugar, en general se reconoce que el modelo de efectos fijos es válido como una prueba de significación de la hipótesis nula general (es decir "no hay efecto en ningún estudio"). Un resultado estadísticamente significativo indica que hay un efecto por lo menos en uno de los estudios.

En tercer lugar, ya sea que haya o no heterogeneidad, el resultado global derivado de un modelo de efectos fijos es una medida promedio del efecto del tratamiento en los estudios en el análisis.

Cuarto, el modelo de efectos aleatorios se basa en la suposición de que los estudios sean una muestra al azar de una población hipotética de estudios y que la heterogeneidad entre los estudios puede ser representada mediante una sola varianza. Además un asunto de tipo pragmático es que el modelo de efectos aleatorios le otorga más peso a los resultados de estudios más pequeños que el modelo de efectos fijos, y tales estudios a menudo tienen una calidad más pobre y pueden ser más susceptibles a estar afectados por sesgos de publicación que los estudios más grandes.

En la práctica, los resultados pueden analizarse mediante un modelo de efectos fijos o un modelo de efectos aleatorios. Si esto se realiza, es mejor decidir con antelación de qué manera serán reportados los resultados para evitar sesgos al momento de decidir en cual modelo se hará énfasis. El modelo de efectos aleatorios tenderá a proporcionar una estimación más conservadora (es decir, con intervalos de confianza más amplios), pero los resultados de los dos modelos deben coincidir básicamente cuando no hay heterogeneidad. Si la prueba de heterogeneidad es estadísticamente significativa y las diferencias en los resultados tienen una importancia práctica, entonces debe enfatizarse más en el modelo de efectos aleatorios. Sin embargo, los revisores deben ser cautelosos en cuanto a la sobreinterpretación de pequeñas diferencias en los intervalos de confianza a partir de análisis diferentes.

Además de los modelos alternativos (de efectos fijos *versus* efectos aleatorios), existen métodos diferentes para los análisis, los cuales pueden utilizarse cuando un revisor desea hacer exploraciones con la ayuda de un estadístico. En general, no existen muchas diferencias entre los resultados derivados de diversos modelos de efectos fijos (Thompson 1991). Hay un menor conocimiento acerca de lo que sucede con diversos modelos de efectos aleatorios.

8.4.5. Resumen

Aunque es deseable mantener la uniformidad en las revisiones con respecto al uso de métodos estadísticos y la presentación de los resultados, es difícil justificar la restricción de los análisis a cualquier método específico. Sin embargo, para ayudar a evitar los sesgos y el mal uso de los métodos estadísticos cada CRG debe elaborar criterios respecto a los métodos estadísticos que va a utilizar. Si un revisor opta por apartarse de los criterios del CRG, deben informarse las razones en SU revisión.

8.5. Presentación de los resultados

En la sección de Métodos deben describirse los métodos analíticos empleados en una revisión, y en la sección de Resultados deben describirse los resultados. Los revisores deben garantizar que la sección de Resultados se concentra en los objetivos y las comparaciones establecidas en el protocolo de la revisión. Aquellos análisis *a posteriori* deben identificarse como tales.

Cuando las revisiones Cochrane se publican, incluyen los datos que los revisores introdujeron en las Tablas de Datos. Las tablas y gráficos que se presentan en RevMan y en la CDSR se generan mediante un programa especial, el MetaView. Éste permite a los revisores y a los usuarios de la CDSR seleccionar métodos estadísticos alternativos y ver los resultados en otras formas distintas a las establecidas en la sección de Métodos de la revisión. Para evitar confusiones, los revisores especifican las presentaciones por defecto para cada resultado en cada revisión. Esto garantiza que los resultados que se presentan son consistentes con lo que se describe en el texto, a menos que alguien que utilice activamente la CDSR decida omitir este punto.

La convención que se utiliza en la CDSR es que si los *odds ratio* o los riesgos relativos para resultados dicotómicos desfavorables son menores que uno (y si las diferencias de riesgos son menores que cero), esto indica que el tratamiento es mejor que el control. Esto es representado en los gráficos mediante estimaciones a la izquierda de la línea vertical, indicando que no hay diferencias. Tanto los revisores como los usuarios pueden confundirse cuando no se mantiene esta convención, en cuyo caso hay un peligro mayor de cometer errores. Sin embargo, existen situaciones en las que puede no ser apropiado adherirse a esta convención de mostrar los resultados "malos". A veces tiene más sentido presentar los resultados para los "buenos" resultados; por ejemplo, nacidos vivos después de un tratamiento para la infertilidad en vez de "no nacidos vivos", o abandono del hábito de fumar después de una intervención para ayudar a las personas a dejar de fuman en vez de "fracaso en abandonar el hábito de fumar". Para algunos resultados, la decisión acerca de si un resultado es "bueno" o "malo" puede reflejar los valores personales. En situaciones como estas, los revisores pueden decidir no adherirse a la convención de reportar los resultados para los desenlaces "malos". Además, en algunos análisis podrían compararse dos intervenciones diferentes, de manera que la decisión acerca de cuál es el "tratamiento" y cual es el "control" es arbitraria. Para acomodarse a estas circunstancias, RevMan 4.0 permite a los revisores que especifiquen los rótulos utilizados para los grupos de "tratamiento" y de "control" en cada resultado. Estos rótulos se utilizan en la CDSR.

8.6. Evaluación de la heterogeneidad

Es importarse preguntarse si los resultados de los estudios son similares dentro de cada comparación o dicho de otra manera: ¿Las diferencias entre los resultados de los estudios son mayores de lo que podría esperarse por azar?. Una manera de responder a ello es examinar una presentación gráfica de los resultados. Si los intervalos de confianza para los resultados de cada estudio (representados típicamente con líneas horizontales) no se sobreponen, esto sugiere que es probable que las diferencias sean estadísticamente significativas (Walker 1988). Las pruebas de homogeneidad son análisis estadísticos formales para examinar si la variación observada en los resultados del estudio es compatible con la variación esperada únicamente por azar. Mientras más significativos sean los resultados de la prueba (el valor p sea más pequeño), es más probable que las diferencias observadas no sean debidas únicamente al azar.

Cuando hay heterogeneidad estadísticamente significativa (una probabilidad baja, tal vez de 1 en 10 o de 1 en 20), esto sugiere que las diferencias observadas en los resultados de los estudios individuales podrían ser causadas por otros factores distintos al azar. Sin embargo, al examinarse las posibles explicaciones para las diferencias, es importante ser cauteloso. Habitualmente los estudios difieren de muchas maneras, y puede ser incorrecto atribuir una diferencia en los resultados a un solo factor.

RevMan evaluará automáticamente la homogeneidad de los resultados de los estudios individuales que se están combinando para cada comparación de datos dicotómicos, continuos o de datos individuales de pacientes. Sin embargo, debe tenerse en cuenta en todas las revisiones, inclusive en aquellas que han utilizado "Otros datos", fundamentos de tipo biológico, psicológico y sociológico cuando se consideran las diferencias entre los estudios.

8.7 Análisis de Subgrupos

Independientemente de la magnitud de la heterogeneidad en los resultados, a menudo resulta interesante examinar una categoría particular de participantes en una revisión (como mujeres, un grupo de edad en particular, o aquellos sujetos con un subtipo específico de enfermedad). Estas evaluaciones, o análisis de subgrupos, son excesivamente frecuentes pero también a menudo son engañosas (Yusuf 1991, Oxman 1992). Las conclusiones que se basan en análisis de subgrupos pueden ser nocivas cuando hacen que se niegue el acceso a un tratamiento efectivo a una categoría particular de personas (una conclusión de tipo "falso negativo"), y cuando hacen que se proporcione un tratamiento que no es efectivo o que inclusive puede ser nocivo, a una categoría particular de personas (una conclusión de tipo "falso positivo"). Los análisis de subgrupos también pueden generar recomendaciones engañosas acerca de las direcciones que deben seguir los estudios en el futuro, que si son adoptadas, podrían malgastar recursos ya escasos.

Debido a estos riesgos y a la frecuencia de su ocurrencia, los revisores necesitan ser cuidadosos al hacer los análisis de subgrupos y al interpretar los resultados de aquellos que ellos consideran que tienen que llevar a cabo. Las siguientes guías pueden utilizarse para ayudar a decidir si se hace un análisis de subgrupos, si una diferencia aparente entre el efecto de la intervención en dichos subgrupos es real y, en consecuencia, cuándo incluir en la revisión una conclusión basada en un análisis de subgrupos (Oxman 1992). También pueden utilizarse para guiar las decisiones acerca de si las diferencias aparentes entre el efecto de diferentes formulaciones de la intervención (tales como la droga o la dosis específica que fueron empleadas) son reales.

- ¿Hay evidencia indirecta que apoye la existencia de una diferencia?

Otra manera de plantear esta pregunta es: ¿Es plausible esperar que habrá una diferencia con base en lo que se conoce acerca de la intervención y del problema? Si no es así, probablemente no debe hacerse el análisis de subgrupos.

- ¿Las hipótesis acerca de la diferencia preceden el análisis y no surgen después de este?
- ¿El análisis de subgrupos es uno entre un pequeño número de hipótesis evaluadas?

En la medida de lo posible, los revisores deben intentar plantear cualquier hipótesis que tengan acerca de diferencias potencialmente importantes en la respuesta de

ciertos subgrupos dentro del protocolo de la revisión, y mantener al mínimo el número de análisis de subgrupos que se han planeado. Mientras más grande sea el número de hipótesis a evaluar, mayor será el número de las diferencias que uno encontrará simplemente por azar. Es más probable que las decisiones acerca de cuáles análisis realizar y cuáles informar estén orientadas por los datos cuando se planean los análisis *a posteriori*, y en consecuencia, es más probable que sean espúreas. De otra parte, cuando una hipótesis ha surgido clara e inequívocamente a partir de un conjunto de datos diferente, el análisis de subgrupo tiene bases más sólidas.

- ¿La diferencia es sugerida a partir de comparaciones dentro de los estudios y no entre los estudios?
- ¿La diferencia es consistente a lo largo de los estudios?

Los revisores deben ser extremadamente cautelosos al examinar y plantear conclusiones acerca de la existencia de diferencias en la respuesta de los subgrupos con base en diferencias en los resultados entre los estudios. Puede haber muchos otros factores aparte de aquel sobre el cual se basan las inferencias, que podrían explicar la diferencia en los resultados entre los estudios. Por ejemplo, diferencias en las drogas, en las dosis, en los esquemas de tratamiento específicos que se han utilizado; en los pacientes (cuyo riesgo de presentar resultados adversos puede variar); en la magnitud de las cointervenciones, o en las maneras de evaluar el resultado, podrían explicar las diferencias entre los resultados de estudios individuales. Para decirlo de una manera sencilla, las inferencias que se basan en comparaciones entre estudios se fundamentan en comparaciones entre grupos que no son comparables. Inclusive si la revisión solamente incluye ensayos aleatorios, debe recordarse que los pacientes no fueron aleatorizados a un estudio o a otro. Las diferencias en la respuesta observada de los subgrupos dentro de un estudio individual proporcionan una base más sólida para plantear conclusiones, y si se observa una diferencia a lo largo de varios estudios de alta calidad, se fortalece la conclusión de que la diferencia es real.

- ¿La magnitud de la diferencia es prácticamente importante?

Si no es así, no importa si las recomendaciones se basan en el análisis de subgrupos o en el análisis global.

- ¿La diferencia es estadísticamente significativa?

RevMan 4.0 no proporciona un método adecuado para evaluar este punto. Dentro de las técnicas estadísticas para llevar a cabo un análisis de subgrupos se incluyen la técnica de Breslow-Day y los modelos de regresión. Con la técnica de Breslow-Day y con otros abordajes similares, es posible utilizar una prueba de heterogeneidad para calcular la probabilidad de que una interacción observada pueda haber surgido por azar. Sin embargo, como se mencionó anteriormente, no se promueve la realización de análisis de subgrupos basados en diferencias entre los estudios.

Con frecuencia se hacen diversas comparaciones de subgrupos diferentes sin que haya una evaluación formal de las interacciones. Cuando se combina esta práctica con el reporte únicamente de aquellos análisis de subgrupos en los cuales se encuentran diferencias considerables de los tratamientos, pueden sobrestimarse la significación y la magnitud de la diferencia. Esto debe evitarse.

Los modelos de regresión, como la regresión logística (Berlin 1994) pueden utilizarse para el análisis de las interacciones, si estas son modeladas de forma multiplicativa. Esto permite evaluar la significación de una interacción, mientras se controla por otros factores. Sin embargo, si hay muchos factores de subgrupo, el número de términos multiplicativos que son necesarios para un modelamiento adecuado de las interacciones puede ser más alto que el número de observaciones, y hacer imposible

un análisis de las interacciones. Un problema adicional se refiere a la decisión de qué términos de interacción dentro de muchos posibles, hay que introducir en el modelo, y a la posibilidad de que haya sesgos en su selección.

Aquellos revisores que no tienen una formación estadística y que desean evaluar formalmente la significación de una diferencia aparentemente importante dentro de un análisis de subgrupos, deben considerar la opción de pedir ayuda a un estadístico. No es importante para los revisores entender los detalles de los abordajes estadísticos que puedan utilizarse, pero es importante entender los conceptos de significación y potencia estadística en el análisis de subgrupos. El análisis estadístico es una herramienta útil para evaluar si una diferencia observada en la respuesta de un subgrupo pudiese haber ocurrido únicamente por azar, pero esto no sustituye al buen juicio, y los métodos estadísticos que se utilicen deben ser adecuados.

8. 8. Métodos estadísticos no disponibles en el RevMan

En algunas revisiones puede ser procedente realizar el análisis sin el apoyo del programa RevMan. Cuando se haga un análisis así, debe describirse y referenciarse en la sección de Métodos de la revisión y sus resultados deben figurar en la sección de Resultados. Los revisores que quieran utilizar tablas para ello, pueden hacerlo con la opción de Tablas Adicionales en RevMan 4.0. En la actualidad no hay un mecanismo para publicar gráficos adicionales en la CDSR, pero estos pueden estar disponibles en otro formato, como por ejemplo en una pà gina en Internet.

8. 8. 1. Metarregresión

La metarregresión puede usarse para evaluar estadísticamente si factores específicos (covariables) influyen en la magnitud del efecto de los estudios (Berlin 1994). Sería preferible que estos análisis se usaran como instrumentos para generar hipótesis. El número de factores que se pueden incluir en un análisis es limitado y existe un sesgo potencial al seleccionar qué factores incluir entre los muchos posibles.

Las covariables de interés pueden ser comunes o no para todos los pacientes en cada estudio. En el caso de características de los estudios que no sean comunes para todos los pacientes (por ejemplo, la media de edad de los pacientes en cada estudio), es importante ser cauteloso con su interpretación debido al riesgo de falacia ecológica (Morgenstern 1982, Greenland 1994). Las covariables que se investigan pueden ser generales (por ejemplo el año del estudio), o específicas de un problema (un subtipo particular de enfermedad). Para la metarregresión pueden usarse diversos métodos estadísticos que incluyen mínimos cuadrados ponderados y modelos logísticos y jerárquicos (Morris 1992, Berlin 1994, McIntosh 1995).

8. 8. 2. Metanálisis bayesiano

Los métodos bayesianos para metanálisis - como el método del perfil de confianza (Eddy 1990, Eddy 1992), (Confidence Profile Method-CPM) - se basan en el principio de que cada observación o conjunto de observaciones debe considerarse conjuntamente con una probabilidad previa que describa un conocimiento ya existente acerca del fenómeno que se analiza. Las nuevas observaciones alteran esta probabilidad para generar una probabilidad posterior. En términos bayesianos, los métodos estadísticos tradicionales usados en metanálisis suponen que la distribución de la probabilidad previa es uniforme, en donde todos los resultados son igualmente

probables. Los modelos bayesianos también permiten la incorporación de evidencia indirecta y opiniones en la generación de las distribuciones de probabilidad previas. Estos enfoques son polémicos porque dependen de las opiniones y a menudo, éstas varían considerablemente. Para realizar análisis bayesianos creíbles, es necesario un desarrollo adicional para la obtención de las distribuciones probabilísticas previas y la realización de análisis robustos (Louis 1993).

8. 8. 3. Métodos " exactos"

El análisis de tablas de 2 X 2 con ceros en algunas casillas (ningún suceso) presentan problemas. Una manera de tratar este problema es agregar una fracción, como ½ , a cada casilla de la tabla. Así se hace en el programa RevMan 4.0. Seguramente es mejor usar cálculos exactos. Hay métodos de cálculo disponibles para calcular un intervalo de confianza exacto para el *odds ratio* (Mosteller 1992) y éstos están disponibles en los paquetes de *software* comercial, por ejemplo, el programa Egret (Egret 1996).

8.9 Otros tipos de datos

8.9.1 Análisis de tiempo transcurrido hasta un evento (supervivencia)

Para las revisiones en las cuales es particularmente importante el tiempo de supervivencia o el tiempo transcurrido hasta un evento, los datos individuales de pacientes son altamente deseables (ver sección 11). Cuando esto no resulta práctico, puede ser posible obtener y analizar los datos derivados de curvas de supervivencia que estén publicadas o calcular un promedio ponderado de la duración media de la supervivencia a lo largo de los estudios (Parmar 1988) (para un ejemplo, ver Grilli 1993). De manera alternativa, pueden hacerse análisis para diferentes periodos de seguimiento utilizando los diferentes puntos en el tiempo como parte del resultado, lo cual se aproxima a un análisis de supervivencia (para un ejemplo, ver Pharoah 1999).

En lo posible, el análisis debe estar basado en datos de supervivencia de participantes individuales. Este abordaje no solo es más robusto (Buyse 1987), sino que también proporciona un conocimiento más profundo acerca del curso de la enfermedad y del efecto del tratamiento a lo largo del tiempo (Clarke 1998). Cuando no es posible obtener estos datos de supervivencia, resulta deseable examinar las distribuciones de los tiempos de supervivencia en aquellos estudios que cuentan con datos individuales disponibles.

8.9.2. Datos ordinales

EN CONSTRUCCION. Se está preparando una nueva sección acerca de los métodos para combinar estudios con datos ordinales (categorías ordenadas que no tienen una distancia natural (numérica) entre las categorías posibles). Una solución que es simple pero que no es totalmente satisfactoria consiste en tratar las escalas ordinales como si fueran escalas continuas. Una solución más complicada pero más sólida consiste en utilizar modelos de *odds* proporcionales (Whitehead 1994).

8.10 Análisis de sensibilidad

Dado que hay diferentes enfoques para realizar una revisión sistemática, los revisores deben preguntarse: "¿Cuán sensibles son los resultados del análisis a los cambios en la manera en que fue realizado?". Esto proporciona a los revisores un mecanismo para probar cuán robustos son los resultados de la revisión con relación a las decisiones y a las asunciones clave surgidas en el proceso de realizar la revisión. Cada revisor debe identificar cómo podría ser posible que las decisiones y las asunciones clave hayan afectado los resultados de una revisión específica. En general, los tipos de decisiones y asunciones que se pueden examinar en los análisis de sensibilidad, incluyen:

- Cambio de los criterios de inclusión (para los tipos de estudios (como por ejemplo, utilizar diferentes puntos de corte metodológicos), los participantes, las intervenciones o medidas de resultado).
- Inclusión o exclusión de estudios donde haya alguna ambigüedad sobre si reúnen los requisitos de inclusión o no.
- Volver a analizar los datos utilizando un intervalo de resultados razonable para los estudios cuyos resultados plantean incertidumbre (por ejemplo, debido a incongruencias (en la manera de informar los resultados) que no puedan resolverse mediante contacto con los investigadores, o debido a diferencias en la definición o medida de los resultados).
- Volver a analizar los datos imputando un intervalo de valores razonable a los datos faltantes.
- Volver a analizar los datos utilizando abordajes estadísticos diferentes (por ejemplo, usar un modelo de efectos aleatorios en vez de un modelo de efectos fijos o viceversa).

Las mismas precauciones que se discutieron para los análisis de subgrupos, se aplican a los análisis de sensibilidad. En particular, dado que muchos análisis de sensibilidad incluyen comparaciones de subgrupos entre los estudios, estos hallazgos necesitan ser interpretados con mucho cuidado.

Si los análisis de sensibilidad que se hacen no cambian sustancialmente los resultados, se fortalece la confianza que pueda atribuirse a tales resultados. Si los resultados cambian de manera que conducen a conclusiones diferentes, esto indica que se necesita tener más cuidado en la interpretación de los resultados y al plantear las conclusiones. Tales diferencias también pueden ayudar a los revisores a clarificar la fuente de las controversias que existen acerca de la efectividad de una intervención, o llevarlos a plantear hipótesis acerca de factores potencialmente importantes que podrían estar relacionados con la efectividad de la intervención y que justifican investigaciones adicionales.

8.11 Temas especiales

8.11.1 Sesgo de publicación y gráficos de embudo (*funnel plots*)

Como se discutió en la sección 5, un componente particularmente importante de una revisión es la identificación de estudios relevantes. El sesgo de publicación ha sido bastante reconocido como un problema en este sentido, ya que quiere decir que la probabilidad de encontrar estudios se relaciona con los resultados de dichos estudios (Begg 1988, Begg 1989, Easterbrook 1991, Dickersin 1992b). Una manera de investigar si una revisión está afectada por un sesgo de publicación consiste en preparar un "gráfico de embudo" (*funnel plot*) y examinar si tiene signos de asimetría. RevMan 4.0 incluye un comando que produce este gráfico. Sin embargo, si hay asimetría, deben considerarse también otras razones distintas al sesgo de publicación.

Los gráficos de embudo se utilizaron por primera vez en investigaciones sobre educación y en la psicología (Light 1984a). Son simples gráficos de dispersión (*scatter plots*) acerca de los efectos de tratamiento calculados a partir de estudios individuales (en el eje X) contra alguna medida del tamaño de muestra de cada estudio (eje Y). La denominación de "gráfico de embudo" surge del hecho de que la precisión en el cálculo del verdadero efecto del tratamiento aumenta a medida que aumenta el tamaño de muestra de los estudios incluidos. Por lo tanto, las estimaciones del efecto que se derivan de estudios pequeños se dispersan ampliamente en la parte inferior del gráfico, mientras que la dispersión es menor entre los estudios más grandes. En ausencia de sesgos, el gráfico se asemeja a un embudo simétrico invertido (ver panel A de la figura).

Las medidas relativas del efecto del tratamiento (como riesgos relativos y *odds ratios*) se representan en una escala logarítmica. Esto garantiza que aquellos efectos de igual magnitud pero con direcciones opuestas (por ejemplo, riesgos relativos de 0.5 y de 2) son equidistantes de 1.0 (Galbraith 1988). Los efectos del tratamiento generalmente se representan contra los tamaños de muestra. Sin embargo, el poder estadístico de un ensayo se define de acuerdo con su tamaño total de muestra y del número de participantes que tuvieron el evento de interés. Por ejemplo, es menos probable que un estudio con 100,000 participantes y con 10 eventos muestre un efecto del tratamiento que sea estadísticamente significativo, que un estudio con 1000 pacientes y 100 eventos. En consecuencia, el error estándar (ES) y la varianza del estimador del efecto se vienen utilizando cada vez más para el eje Y en los gráficos de embudo que el tamaño total de la muestra. RevMan 4.0 utiliza el valor de 1/SE, representado contra el tamaño del efecto calculado mediante el método estadístico seleccionado por el revisor para el resultado particular.

Si hay un sesgo, por ejemplo, debido a que dejan de publicarse estudios más pequeños que no tienen efectos estadísticamente significativos (que se muestran como círculos abiertos en la figura), esto conducirá a una apariencia asimétrica del gráfico de embudo, con un hueco en la esquina inferior del gráfico (panel B). En esta situación, el efecto calculado en un metanálisis sobrestimará el efecto del tratamiento (Villar 1997, Egger 1997b). Mientras más pronunciada sea la asimetría, es más probable que la magnitud del sesgo sea considerable.

El sesgo de publicación ha sido bastante asociado con una asimetría en el gráfico de embudo (Light 1984a). Sin embargo, el gráfico de embudo debe ser considerado como un recurso genérico para examinar si los estudios más pequeños en un metanálisis tienden a mostrar efectos de tratamiento más grandes, lo cual puede ser producto de otras razones distintas al sesgo de publicación (Egger 1997a, Egger 1998b). Algunas de ellas se muestran en la tabla:

Posibles fuentes de asimetría en los gráficos de embudo

1. Sesgos de selección
    - Sesgo de publicación
    - Sesgo de identificación
    - Sesgo de lenguaje
    - Sesgo de citación
    - Sesgo de múltiples publicaciones
2. Baja calidad metodológica de estudios más pequeños
    - Diseño metodológico deficiente
    - Análisis inadecuado
    - Fraude
3. Heterogeneidad verdadera

El tamaño del efecto difiere de acuerdo con el tamaño del estudio (por ejemplo, debido a diferencias en la intensidad de las intervenciones o a diferencias en el riesgo subyacente entre estudios de diferente tamaño).

4. De artefacto
5. Por azar.

Aun cuando un estudio haya sido publicado, la probabilidad de identificarlo también está influenciada por sus resultados. Por ejemplo, el sesgo de lenguaje (preferir la publicación de estudios que no tienen hallazgos significativos en otros idiomas diferentes al inglés) hace menos probable que se encuentren dichos estudios "negativos" (Grégoire 1995, Egger 1997c). El sesgo de citación hace que los estudios "negativos" sean referenciados con menos frecuencia y que por lo tanto sea más probable que no se identifiquen cuando se hacen búsquedas de ensayos relevantes (Gotzsche 1987, Gotzsche 1989, Ravnskov 1992). Por el contrario, los resultados de ensayos "positivos" a veces son publicados más de una vez, lo cual alimenta la probabilidad de que sean identificados (sesgo de publicación múltiple) (Gotzsche 1989, Huston 1996, Tramèr 1997).

Otra fuente de asimetría surge a partir de diferencias en la calidad metodológica. Los estudios más pequeños, en promedio se realizan y se analizan con menos rigor metodológico que los estudios más grandes. Los ensayos de baja calidad también tienden a mostrar efectos del tratamiento más grandes (Schulz 1995, Moher 1998a). Si hubieran sido adecuadamente realizados y analizados, algunos ensayos que serían "negativos" hubieran podido convertirse en "positivos" (panel C).

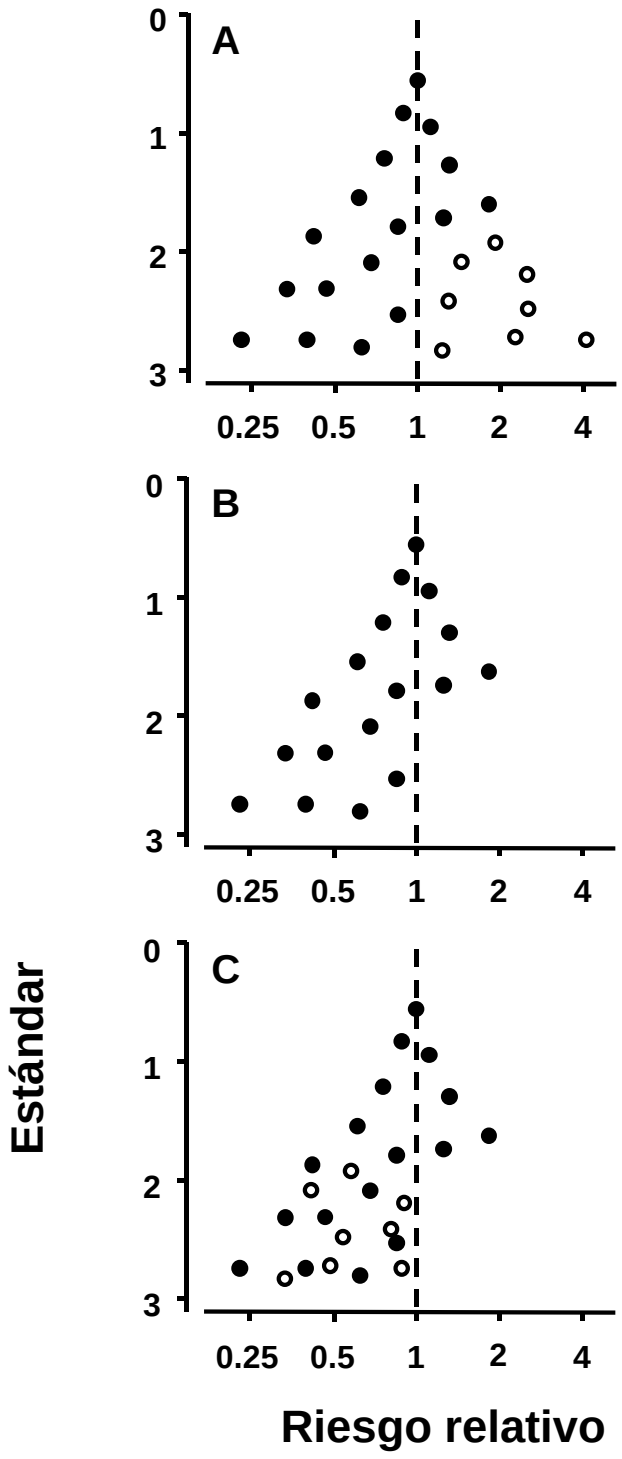
Una heterogeneidad verdadera en los efectos del tratamiento también puede explicar la asimetría del gráfico de embudo. Por ejemplo, puede observarse un beneficio sustancial solamente en aquellos pacientes en alto riesgo de presentar el resultado que es afectado por la intervención, y estos pacientes de alto riesgo habitualmente tienen una probabilidad mayor de ser incluidos en estudios pequeños y precoces (Davey Smith 1994, Glasziou 1995). Adicionalmente, los ensayos pequeños tienden a realizarse antes de que se inicien ensayos grandes y durante los años intermedios el tratamiento convencional puede haber mejorado. Además, algunas intervenciones pueden haberse implementado con menos cuidado en los ensayos más grandes, y por ende haber provocado estimaciones menores del efecto del tratamiento (Stuck 1998). También se ha planteado que la asimetría del gráfico de embudo puede ser un artefacto (Irwig 1998), pero los estudios de simulación han demostrado que esto ocurre con poca frecuencia, cuando el efecto global del tratamiento es muy grande y cuando el resultado de interés es raro (Sterne, sin publicar). Finalmente, por supuesto, es posible que un gráfico de embudo asimétrico surja simplemente por azar.

La simetría o la asimetría se define generalmente de manera informal, a través de un examen visual, pero la interpretación visual de los gráficos de embudo puede variar entre los observadores (Villar 1997). Se han propuesto métodos estadísticos más formales para examinar las asociaciones entre los efectos del estudio y su tamaño (Begg 1994, Egger 1997b). En la actualidad hay debates acerca de las propiedades estadísticas, las ventajas y limitaciones de estas pruebas (Naylor 1997, Irwig 1998, Seagrott 1998, Egger 1998b). Dichas pruebas no se encuentran disponibles en RevMan 4.0. Actualmente se llevan a cabo trabajos metodológicos que examinan estos aspectos, pero resulta claro que el examen visual y el análisis estadístico de los gráficos de embudo tienen un poder limitado para detectar sesgos si el número de estudios es bajo.

Los revisores deben examinar el gráfico de embudo relevante cada vez que hacen un metanálisis. Si hay asimetría, deben explorarse sus posibles causas. Sin embargo, el

poder de este método es más limitado en aquellas situaciones en donde es más probable que un sesgo distorsione los resultados de los metanálisis: cuando incluye solamente unos cuantos estudios pequeños. Finalmente, debe recordarse que aunque los gráficos de embudo pueden alertar a los revisores acerca de un problema que necesita ser considerado, no proporcionan una solución para esto. La única forma satisfactoria de abordar el sesgo de información y la inadecuada calidad de ensayos individuales es mediante un registro prospectivo de los ensayos (Simes 1986, Dickersin 1988, Anonymous 1991, Dickersin 1992a) y mejorando la calidad de la conducción, el análisis y el reporte de estudios, metanálisis y revisiones sistemáticas (Begg 1996, Moher 1995b).

Inscripción de la figura: Gráficos de embudo hipotéticos. Panel A: gráfico simétrico libre de sesgos; Panel B: gráfico asimétrico en presencia de un sesgo de información, Panel C: gráfico asimétrico en presencia de un sesgo debido a baja calidad metodológica de los estudios más pequeños.



8.11.2 Errores de la unidad de análisis (ensayos aleatorios agrupados)

En algunos estudios, las personas son asignadas a intervenciones diferentes según agrupamientos (por ejemplo, centros hospitales, o comunidades) para evitar contaminaciones o bien por conveniencia. A menudo, cuando esto se hace, la unidad de asignación es diferente de la unidad de análisis, en donde las personas son asignadas de acuerdo con grupos, pero se analizan como si hubieran sido asignadas

individualmente. Esto habitualmente se denomina “error de la unidad de análisis” (Whiting-O´Keefe 1984). Efectivamente, cuando se utilizan los individuos como la unidad de análisis cuando se han asignado grupos de personas, se aumenta la potencia de los estudios, al incrementar los grados de libertad. Esto puede resultar en conclusiones de tipo falso positivo que indiquen que la intervención tuvo un efecto cuando en realidad no es así. En el contexto de una revisión, esto puede hacer que los estudios tengan intervalos de confianza más estrechos y que reciban más peso que lo que sería apropiado. Esta situación también puede presentarse si las personas son asignadas a intervenciones que después se aplican a partes de ellas (como ojos, dientes o juanetes), pero se analizan las partes separadas y no las personas.

Los revisores deben obtener información acerca de la unidad de aleatorización y la unidad de análisis cuando esto sea relevante. Para evitar un error de la unidad de análisis, los análisis deben realizarse al mismo nivel que la asignación, pero esto podría reducir considerablemente y de manera innecesaria el poder del estudio. Si es posible, debe obtenerse una estimación de la variabilidad relativa al interior de los agrupamientos y entre ellos (el “coeficiente de correlación intraclase”) y posiblemente utilizarlo para ajustar los resultados (Cornfield 1978, Donner 1981).

8.11.3 Comparaciones múltiples y papel del azar

Para las revisiones y los estudios que tienen varios análisis de subgrupos y múltiples resultados, si se hace un número suficiente de análisis, es probable que algunos de ellos produzcan resultados “estadísticamente significativos” solo por azar. Por ejemplo, al usarse la prueba convencional para evaluar la significación estadística (P = 0.05), una en 20 comparaciones tiene la probabilidad de ser “estadísticamente significativa” incluso cuando en verdad no hay diferencias entre las intervenciones que se están comparando. Mientras más análisis se hagan, es más probable que se encuentre que algunos sean estadísticamente significativos solamente por azar.

Debido a este problema, la fuerza de la evidencia emanada de una revisión o un estudio depende de lo focalizadas que hayan sido sus preguntas desde el principio. Lamentablemente, cuando se presentan los resultados de un estudio, no siempre es posible saber cuántas comparaciones se hicieron. Es probable que los hallazgos interesantes hayan sido seleccionados para su presentación o publicación a partir de un gran número que no son nada interesantes, lo que lleva a conclusiones espúreas.

Los revisores deberían distinguir, si fuera posible, entre datos que no se recogieron en un estudio y datos recogidos pero no presentados. Esto es particularmente importante en los estudios que informan sobre resultados múltiples, debido a la posibilidad de informes sesgados (Gotzsche 1989, Gotzsche 1990).

No hay ninguna solución sencilla ni totalmente satisfactoria para el problema de las comparaciones múltiples, pero pueden ofrecerse los siguientes consejos:

- Limitar al mínimo los análisis de subgrupos
- Mencionar *a priori* en el protocolo de la revisión qué análisis y resultados son de interés particular (cuantos menos, mejor).
- Aunque se recomienda que deben informarse en las revisiones Cochrane todos los resultados de los estudios incluidos que probablemente pueden ser importantes para las personas, la fuerza de la evidencia es menor si hay múltiples comparaciones, particularmente para aquellos resultados que no fueron preestablecidos.

- Interpretar con cautela cualquier hallazgo que no fue contemplado en las hipótesis, incluso cuando sea “estadísticamente significativo”.

8.11.4. Datos faltantes

Pueden perderse tres tipos de datos en una revisión: estudios no identificados, información para calcular la magnitud de los efectos e información acerca de ciertas características que pueden estar relacionadas con la magnitud de los efectos. El primer tipo de datos faltantes se discute en la sección 8.11.1 relacionada con el sesgo de publicación. Dentro de los métodos para manejar la información perdida dentro de los estudios que son identificados, se incluye la simple exclusión de aquellos estudios que tienen datos faltantes, imputaciones o estrategias complejas para asumir valores para los datos faltantes (Follmann 1992, Pigott 1993). No existen soluciones simples para el problema de los datos faltantes. En general:

- En lo posible, deben contactarse los investigadores para garantizar que no falten datos para sus estudios.
- Deben hacerse explícitas las suposiciones de cualquier método que se utilice para abordar el problema de los datos faltantes.
- Deben hacerse análisis para evaluar qué tan sensibles son los resultados ante cambios razonables en las suposiciones que se hacen (ver sección 8.10).

8.11.4 Ensayos cruzados

El análisis de ensayos cruzados es problemático, particularmente si se incluyen simultáneamente ensayos con grupos paralelos y con diseño cruzado. Una estrategia consiste en analizar los ensayos cruzados de la misma manera que los ensayos con grupos paralelos. Esto es, se utilizan los resultados del periodo de tratamiento como si fueran provenientes de un grupo de pacientes, y los resultados del periodo de control se utilizan como si fueran provenientes de un grupo independiente de pacientes. Sin embargo, este abordaje ignora el hecho de que los mismos pacientes estarán en ambos brazos del estudio y que claramente estos no son independientes. Como alternativa, un abordaje más apropiado es emplear técnicas específicamente desarrolladas para diseños pareados y para combinar diseños pareados con diseños no pareados (Duffy 1989). Si ambos abordajes llegan a resultados similares, los resultados de la primera estrategia (la cual puede ser realizada en RevMan 4.0) tiene bases más sólidas. Para un ejemplo de este abordaje, ver (Jaeschke 1990).

8.11.5 Metanálisis acumulativo

En un metanálisis acumulativo, se añaden los estudios, uno a la vez y guardando un orden establecido (por ejemplo, de acuerdo a su fecha de publicación o a su calidad metodológica) (Lau 1992). Este abordaje puede ser utilizado como un análisis de sensibilidad para determinar si los resultados han sido consistentes a lo largo del tiempo o entre estudios de diferente calidad. Los metanálisis acumulativos pueden hacerse en RevMan 4.0 de-seleccionando todos los estudios del análisis y luego seleccionándolos en un orden específico y ejecutando un análisis cuando sea necesario. Sin embargo, no es posible reproducir estos análisis con facilidad cuando la revisión se publica en la CDSR. Si se va a hacer un metanálisis acumulativo, esto debe mencionarse en el protocolo de la revisión.

8.11.6 Análisis de dosis-respuesta

Los mismos principios descritos para los análisis de subgrupos, pueden ser aplicados a los análisis de dosis-respuesta y de duración de la respuesta. Pueden utilizarse métodos de metarregresión para calcular los parámetros de dosis-respuesta (Greenland 1992, Berlin 1993). Las conclusiones acerca de las diferencias en el efecto debido a diferencias en las dosis tienen una base más sólida si se aleatoriza a los participantes a una u otra dosis dentro de un estudio y si se encuentra una relación consistente a lo largo de estudios similares. Aunque los revisores deben tener en cuenta los efectos de las dosis como una posible explicación para la heterogeneidad, deben tener cuidado en plantear conclusiones acerca de los efectos de la dosis o de la duración de los efectos con base en diferencias entre los estudios.

8.12 Dónde pedir ayuda

El primer punto de contacto para los revisores que necesitan apoyo estadístico debe ser el coordinador de su CRG. Cada CRG debe establecer normas acerca de los análisis estadísticos, e idealmente debe haber un estadístico que trabaje con cada CRG. Sin embargo, puede ser más conveniente para algunos revisores consultar un estadístico local. Si un revisor necesita apoyo estadístico u otro tipo de ayuda con sus análisis que no pueden encontrar directamente a través del equipo editorial de su CRG o a través de apoyo local, puede pedirse ayuda a un Centro Cochrane. Como norma, las solicitudes de ayuda no deben realizarse directamente al Grupo de Métodos Estadísticos (Statistical Methods Group – SMG). El SMG es responsable de proporcionar formación y asesoría en las normas, pero no cuenta con los recursos necesarios para apoyar directamente a revisores individuales. Los estadísticos que trabajan con los CRG son miembros del SMG y de ser necesario consultarán con otros miembros del Grupo de Métodos. De manera similar, los Centros Cochrane dirigirán las solicitudes al SMG cuando sea apropiado. Sin embargo, para lograr un mejor uso de un recurso que es relativamente escaso, es importante que el SMG no se vea abrumado con asuntos que, por lo menos para los estadísticos, son solicitudes de rutina.

Puede encontrarse ayuda adicional en la guía del usuario de RevMan, en las referencias citadas en esta sección del Manual, en otras referencias en la *Cochrane Review Methodology Database* y en la lista de preguntas frecuentes (FAQ) de RevMan. Si usted descubre que existen vacíos importantes en la ayuda que pueda encontrar en el Manual o en la guía del usuario de RevMan, o bien existen problemas específicos que no son abordados, por favor, háganoslo saber (Mike Clarke o Andy Oxman). Si usted tiene soluciones a problemas que otros pueden encontrar, por favor compártalas con nosotros. El Manual continuará evolucionando en base a las necesidades expresadas por aquellos que lo utilizan, en su experiencia y en los nuevos desarrollos metodológicos, lo cual es especialmente cierto para esta sección de Análisis y Presentación de Resultados.

8.13 Referencias

Altman 1992. Altman DG, Gardner MJ. Confidence intervals for research findings. Br J Obstet Gynaecol 1992; 99:90-1.

Altman 1998. Altman DG. Confidence intervals for the number needed to treat. BMJ 1998; 317:1309-12

Anonymous 1991. Anonymous. Making clinical trialists register. Lancet 1991; 338:244-5.

Antman 1992. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA 1992; 268:240-8.

Begg 1988. Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. J Roy Stat Soc A 1988; 151:419-63.

Begg 1989. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. J Natl Cancer Inst 1989; 81:107-15.

Begg 1994. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994;50:1088-99.

Begg 1996. Begg CB, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637-639.

Berlin 1989. Berlin JA, Begg CB, Louis TA. An assessment of publication bias using a sample of published clinical trials. J Am Stat Assoc 1989; 84:381-92.

Berlin 1993. Berlin JA, Longnecker MP, Greenland S. Meta analysis of epidemiologic dose-response data. Epidemiology 1993; 4:218-28.

Berlin 1994. Berlin JA, Antman EM. Advantages and limitations of meta-analytic regressions of clinical trials data Advantages and limitations of metaanalytic regressions of clinical trials data. Online J Curr Clin Trials [serial online] 1994; Doc No 134.

Buyse 1987. Buyse ME, Ryan LM. Issues of efficiency in combining proportions of deaths from several clinical trials. Stat Med 1987; 6:565-76.

Clarke 1998. Clarke M, Godwin J. Systematic reviews using individual patient data: a map for the minefields? Ann Oncol 1998;9:827-33.

Cooper 1980. Cooper HM, Rosenthal R. Statistical versus traditional procedures for summarizing research findings. Psychol Bull 1980; 87:442-9

Cornfield 1978. Cornfield J. Randomization by group: a formal analysis. Am J Epidemiol 1978; 108:100-2.

Davey Smith 1994. Davey Smith G, Egger M. Who benefits from medical interventions? Treating low risk patients can be a high risk strategy. Br Med J 1994;308:72-4.

Detsky 1992. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta analysis. J Clin Epidemiol 1992; 45:255-65.

Dickersin 1988. Dickersin K. Report from the panel on the Case for Registers of Clinical Trials at the Eighth Annual Meeting of the Society for Clinical Trials. Controlled Clin Trials 1988; 9:76-81.

Dickersin 1992a. Dickersin K. Keeping posted. Why register clinical trials? - revisited. Controlled Clin Trials 1992; 13:170-7.

Dickersin 1992b. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. JAMA 1992; 263:374-8.

Donner 1981. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. Am J Epidemiol 1981; 114:906-14.

Duffy 1989. Duffy SW, Rohan TE, Altman DG. A method for combining matched and unmatched binary data. Application to randomized, controlled trials of photocoagulation in the treatment of diabetic retinopathy. Am J Epidemiol 1989; 130:371-8.

Easterbrook 1991. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991; 337:867-72.

Eddy 1990. Eddy DM, Hasselblad V, Schachter R. An introduction to a Bayesian method for meta-analysis: the confidence profile method. Med Decis Making 1990; 10:15-23.

Eddy 1992. Eddy DM, Hasselblad V, Shachter R. Meta-analysis by the confidence profile method. The statistical synthesis of evidence. Boston: Academic Press, 1992.

Egger 1997a. Egger M, Davey Smith G, Phillips AN. Meta-analysis: principles and procedures. Br Med J 1997;315:1533-7.

Egger 1997b. Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. Br Med J 1997;315:629-634.

Egger 1997c. Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. Lancet 1997;350:326-329.

Egger 1998. Egger M, Davey Smith G. Meta-analysis: Bias in location and selection of studies. Br Med J 1998;316:61-6.

Egger 1998b. Egger M, Davey Smith G, Minder C. Authors' reply. Br Med J 1998;316:471.

Egret 1996. Egret statistical software. Seattle: Statistics in Epidemiology Research Corporation, 1996.

Fleiss 1993. Fleiss JL. Measures of effect size for categorical data. In: Cooper H, Hedges LV, editors. The Handbook Of Research Synthesis. New York: Russell Sage Foundation, 1993; 245-81.

Follmann 1992. Follmann D, Elliott P, Suh I, Cutler J. Variance imputation for overviews of clinical trials with continuous response. J Clin Epidemiol 1992; 45:769-73.

Galbraith 1988. Galbraith R. A note on graphical presentation of estimated odds ratios from several clinical trials. Stat Med 1988;7:889-94.

Glasziou 1995. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. BMJ 1995;311:1356-9.

Gotzsche 1987. Gotzsche PC. Reference bias in reports of drug trials. Br Med J 1987;295:654-6.

Gotzsche 1989. Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis [published erratum appears in Controlled Clin. Trials 1989;50:356]. Controlled Clin Trials 1989; 10:31-56.

Gotzsche 1990. Gotzsche PC. Sensitivity of effect variables in rheumatoid arthritis: a meta-analysis of 130 placebo controlled NSAID trials [published erratum appears in J Clin Epidemiol 1991;44:613]. J Clin Epidemiol 1990; 43:1313-8.

Greenland 1992. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. Am J Epidemiol 1992; 135:1301-9.

Greenland 1994. Greenland S, Robins J. Invited commentary: ecologic studies - biases, misconceptions, and counterexamples. Am J Epidemiol 1994; 139:747-50.

Grégoire 1995. Grégoire G, Derderian F, LeLorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? J Clin Epidemiol 1995;48:159-63.

Grilli 1993. Grilli R, Oxman AD, Julian JA. Chemotherapy for advanced non-small-cell lung cancer: How much benefit is enough? J Clin Oncology 1993; 11:1866-72.

Hedges 1992. Hedges LV. Modelling publication selection effects in meta-analysis. Stat Sci 1992; 7:246-55.

Huston 1996. Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. Lancet 1996;347:1024-6.

Irwig 1998. Irwig L, Macaskill P, Berry G. Graphical test is itself biased [letter]. Br Med J 1998;316:470.

Iyengar 1988. Iyengar S, Greenhouse JB. Selection models and the file drawer problem. Stat Sci 1988; 3:109-35.

Jaeschke 1990. Jaeschke R, Oxman AD, Guyatt GH. To what extent do patients with chronic heart failure and sinus rhythm benefit from digoxin therapy? A systematic overview (meta-analysis). Am J Med 1990; 88:279-86.

Laird 1990. Laird NM, Mosteller F. Some statistical methods for combining experimental results. Int J Tech Assess in Healthcare 1990; 6:5-30.

Lau 1992. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. N Engl J Med 1992; 327:248-54.

Laupacis 1988. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med 1988; 318:1728-33.

Light 1984. Light RJ, Pillemer DB. Summing up. The science of reviewing research. Cambridge, Massachusetts, and London, England: Harvard University Press, 1984.

Louis 1993. Louis T, Zelterman D. Bayesian approaches to research synthesis. In: Cooper H, Hedges LV, editors. The Handbook Of Research Synthesis. New York: Russell Sage Foundation, 1993; 411-22.

McIntosh 1995. McIntosh M. The population risk as an explanatory variable in research synthesis of clinical trials. Harvard Medical School 1995; 1-35.

Moher 1995. Moher D, Olkin I. Meta-analysis of randomised controlled trials. A concern for standards. JAMA 1995;274:1962-3.

Moher 1998. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, Tugwell P, Klassen TP. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998;352:609-13.

Morgenstern 1982. Morgenstern H. Uses of ecological analysis in epidemiologic research. Am J Public Health 1982; 72:1336-44.

Morris 1992. Morris CN, Normand SL. Hierarchical models for combining information and for meta-analyses. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. Bayesian Statistics, 4th edition. New York: Oxford University Press, 1992.

Mosteller 1992. Mosteller F, Chalmers TC. Some progress and problems in meta-analysis of clinical trials. Stat Sci 1992; 7:227-36.

Naylor 1995. Naylor CD. The case for failed meta-analyses. J Evaluation Clin Practice 1995; 1:127-30.

Naylor 1997. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. Br Med J 1997;315 :617-9.

Oxman 1992. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992; 116:78-84.

Parmar 1998. Parmar MK, Torri V, Stewart LA. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 1998;17:2815-34.

Pharoah 1999. Pharoah FM, Mari JJ, Streiner D. Family intervention for schizophrenia (Cochrane Review). In: The Cochrane Library, Issue 2, 1999. Oxford: Update Software.

Pigott 1993. Pigott TD. Methods for handling missing data in research synthesis. In: Cooper H, Hedges LV, editors. The Handbook Of Research Synthesis. New York: Russell Sage Foundation, 1993; 163-75.

Ravnskov 1992. Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. Br Med J 1992;305:15-9.

Schulz 1993. Schulz KF, Altman DG. Statistical methods for data synthesis. Cochrane Workshop Report. Oxford: UK Cochrane Centre, 1993.

Schulz 1995. Schulz KF, Chalmers I, Hayes RJ, Altman D. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408-12.

Seagrott 1998. Seagrott V, Stratton I. Test had 10% false positive rate [letter]. Br Med J 1998;316:470.

Simes 1986. Simes RJ. Publication bias: the case for an international registry of clinical trials. J Clin Oncol 1986; 4:1529-41.

Sinclair 1994. Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. J Clin Epidemiol 1994; 47:881-9.

Smeeth 1999. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses – sometimes informative, usually misleading. Brit Med J 1999;318:1548-51.

Sterne (submitted). Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: Simulation and prevalence studies (submitted).

Stuck 1988. Stuck AE, Rubenstein LZ, Wieland D. Bias in meta-analysis detected by a simple, graphical test. Asymmetry detected in funnel plot was probably due to true heterogeneity [letter]. Br Med J 1998;316:469-71.

Thompson 1991. Thompson SG, Pocock SJ. Can meta-analyses be trusted? Lancet 1991; 338:1127-30.

Tramèr 1997. Tramèr MR, Reynolds DJM, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. Br Med J 1997;315:635-40.

Villar 1997. Villar J, Piaggio G, Carroli G, Donner A. Factors affecting the comparability of meta-analyses and largest trials results in perinatology. J Clin Epidemiol 1997;50:997-1002.

Walker 1988. Walker AM, Martin-Moreno JM, Artalejo FR. Odd man out: a graphical approach to meta-analysis. Am J Public Health 1988; 78:961-6.

Whitehead 1994. Whitehead A, Jones NMB. A meta-analysis of clinical trials involving different classifications of response into ordered categories. Stat Med 1994; 13:2503-15.

Whitehead 1999. Whitehead A, Bailey AJ, Elbourne D. Combining summaries of binary outcomes with those of continuous outcomes in a meta-analysis. J Biopharm Stat 1999;9:1-16.

Whiting-O'Keefe 1984. Whiting-O'Keefe QE, Henke C, Simborg DW. Choosing the correct unit of analysis in medical care experiments. Med Care 1984; 22:1101-14.

Yusuf 1991. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991; 266:93-8.

9. INTERPRETACIÓN DE LOS RESULTADOS

Aunque se puede argumentar que los resultados de una revisión sistemática deben hablar por sí mismos, a la hora de decidir muchas personas buscan en las secciones de discusión y conclusiones de una revisión una ayuda para interpretar sus resultados. En verdad, muchas personas prefieren ir directamente a las conclusiones sin haber examinado antes el resto de la revisión.

Existen algunos temas que deben ser objeto de discusión y conclusiones, y que pueden ser de ayuda en la toma de decisiones:

- La solidez de la evidencia.
- La aplicabilidad de los resultados.
- Otras informaciones, como son las consideraciones de los costes y la práctica actual, que quizá sean relevantes para alguien que debe tomar una decisión.
- Clarificación del balance entre los beneficios esperados, y los posibles daños y costes de la intervención de interés.

Dado que las revisiones Cochrane tienen un público internacional, la discusión y las conclusiones deben, tanto como sea posible, tomar una perspectiva internacional amplia en lugar de restringirse a cuestiones nacionales o locales. Los revisores deben ser particularmente cuidadosos y tener presente que es posible que diferentes personas tomen decisiones distintas basadas en las mismas evidencias. El objetivo primario de la discusión y de las conclusiones debe ser ayudarles a comprender las implicaciones de esta evidencia en relación con las decisiones prácticas. Deben evitarse las recomendaciones que se basen en alguna asunción acerca de las prioridades y los recursos disponibles.

9. 1. La solidez de la evidencia.

Un buen punto de partida para la sección de discusión de una revisión es abordar cualquier limitación metodológica importante de los ensayos incluidos y de los métodos usados en la revisión, que pueda afectar a las decisiones sanitarias prácticas a tomar o en futuras investigaciones. Ésta no debe ser una discusión detallada sobre los estudios o los métodos de la revisión, y la información que se proporcione en la sección de calidad metodológica no debe repetirse aquí.

A menudo es útil discutir cómo los estudios incluidos encajan en el contexto de otras evidencias que no están incluidas en la revisión. Por ejemplo, para las revisiones sobre farmacoterapia quizá sea pertinente referirse a los estudios de dosificación o estudios no aleatorizados sobre el riesgo de efectos adversos poco comunes. Si se citan referencias de otros tipos de evidencia debe mencionarse claramente si fueron revisadas sistemáticamente.

Al considerar cuán consistentes son las pruebas para concluir que existe una relación de causa y efecto entre una intervención y un resultado importante, puede ser de utilidad valorar la evidencia indirecta de una relación. Esto incluye la información relacionada con los resultados intermedios (como son las medidas fisiológicas o los indicadores bioquímicos que son marcadores de riesgo para el resultado de interés), la procedente de los estudios realizados en diferentes poblaciones (incluidos los estudios en animales) y la proveniente de relaciones análogas (como son, por ejemplo, las intervenciones similares).

Dado que las conclusiones acerca de la eficacia de una intervención son esencialmente inferencias causales, tal vez los revisores quieran considerar unas orientaciones para evaluar la fuerza de una inferencia causal, como las propuestas por Hill (34). En el contexto de una revisión sistemática de ensayos clínicos, estas consideraciones incluirían:

- ¿Qué nivel de calidad tienen los ensayos incluidos?
- ¿Cuán amplios y significativos son los efectos observados?
- ¿Cuán consistentes son los efectos entre los ensayos?
- ¿Hay una clara relación dosis-respuesta?
- ¿Existe evidencia indirecta que apoye la inferencia?
- ¿Se han descartado otras explicaciones verosímiles de los efectos observados (por ejemplo sesgo o cointervención)?

Se han descrito métodos más o menos explícitos para calificar el nivel de evidencia subyacente a una conclusión (CTFPHE 1979, Cook 1992, Gyorkos 1994, Guyatt 1995, US PSTF 1996) aunque no hay un único enfoque universalmente aceptado para la amplísima variedad de revisiones incluidas en la Base de Datos Cochrane de Revisiones Sistemáticas. Un CRG tal vez opte por usar un método estándar para calificar la validez de la evidencia en las revisiones incluidas en su módulo y, con el tiempo, quizá se llegue a que la Colaboración desarrolle una metodología más uniforme y explícita para sacar conclusiones acerca de la solidez general de las evidencias en las conclusiones principales de una revisión. Sin embargo, la selección de un método para resumir el grado de evidencia indicado al tema que está revisándose depende de los revisores individuales, en consulta con otros de su Grupo.

9. 2. Aplicabilidad.

"Se necesita un acto de fe siempre que se aplica el resultado de un estudio a la población en general" o a una persona determinada*. "Al dar ese paso, uno siempre debe lograr un equilibrio entre hacer amplias generalizaciones justificables y tomar un punto de vista conservador respecto de las conclusiones" (*Friedman 1985*).*

Los usuarios de las revisiones Cochrane deben decidir, implícita o explícitamente, si la evidencia que encuentran en ellas es aplicable a sus circunstancias específicas o no. Para hacerlo, en primer lugar deben considerar si la revisión contiene información válida sobre beneficios y daños de interés para ellos. En la medida en que esto sea así, entonces necesitan decidir si los participantes y los entornos en los estudios incluidos son razonablemente similares a su propia situación. Además, a menudo será importante para ellos considerar las características de las intervenciones o de la atención adicional proporcionada en los estudios incluidos, para decidir acerca de la aplicabilidad de estas evidencias.

La valoración definitiva sobre la aplicabilidad de la revisión depende de las situaciones específicas para las que han de tomarse las decisiones sanitarias. Al abordar la aplicabilidad de los resultados de una revisión, los revisores deben ser cautelosos para no suponer que sus propias circunstancias, o las circunstancias reflejadas en los estudios incluidos, son las mismas que las de los demás. Los revisores pueden, sin embargo, ayudar a tomar decisiones sobre la aplicabilidad, llamando la atención sobre el espectro de situaciones en las que la evidencia es aplicable de modo procedente, aquellas en las que no lo es y la previsible variación de los efectos en diferentes situaciones.

En general, en lugar de aplicar rígidamente los criterios de inclusión y de exclusión de los estudios a circunstancias específicas, es mejor preguntarse si hay razones imperiosas por las cuales los ensayos no deben aplicarse en ciertas circunstancias (Guyatt 1994, Dans 1996). A veces los revisores pueden ayudar a quienes toman decisiones identificando aquellas fuentes de variación que puedan limitar la aplicabilidad de resultados, tales como:

- Variación biológica y cultural.
- Variación en el cumplimiento del tratamiento.
- Variación en los riesgos de base.

Al abordar estos temas, no puede esperarse que los revisores sean conscientes o valoren los innumerables escenarios que podemos hallar en el mundo. Sin embargo, sí que pueden tratar las diferencias de conocida importancia para mucha gente y, fundamentalmente, deben evitar dar por sentado que las circunstancias de otras personas son las mismas que las suyas propias a la hora de escribir los resultados y sacar conclusiones.

9. 2. 1. Variación biológica y cultural.

Los temas de la variación biológica que deberían considerarse incluyen divergencias en cuanto a la fisiopatología (por ejemplo, diferencias biológicas entre mujeres y hombres que pueden modificar la respuesta a un tratamiento) y la divergencia relacionada con un agente causal (por ejemplo, en enfermedades infecciosas como la malaria). En el caso de ciertos problemas sanitarios -como son los psiquiátricos-, las diferencias culturales a veces pueden limitar la aplicabilidad de los resultados.

9. 2. 2. Variación en el cumplimiento del tratamiento.

La variabilidad en el cumplimiento, tanto de quienes reciben atención sanitaria como de quienes la prestan, puede limitar la aplicabilidad de los resultados de la revisión. Las diferencias en el cumplimiento pueden deberse a diferencias en la condición económica o a actitudes que hacen inaccesibles o impracticables algunas prestaciones sanitarias en determinados entornos como, por ejemplo, en países en vías de desarrollo (Dans 1996).

9. 2. 3. Variación en el riesgo de base.

El beneficio neto de cualquier intervención depende del riesgo de padecer un resultado adverso sin la intervención, así como de la eficacia de ésta. Por consiguiente, la variación en el riesgo de base es siempre una consideración importante para determinar la aplicabilidad de los resultados. Sin embargo, es importante distinguir dos aspectos. El primero es si son aplicables los beneficios y los riesgos descritos. Por ejemplo, puede dudarse que los resultados obtenidos en pacientes de alto riesgo sean aplicables a pacientes de bajo riesgo, o si son aplicables a pacientes con comorbilidad. Si no hay una razón imperiosa para dar por sentado que los beneficios y riesgos sí son directamente aplicables, es posible estimar la repercusión esperada de una intervención (por ejemplo el número de personas que es necesario tratar-NNT) aplicando el efecto relativo estimado de una intervención (por ejemplo, el riesgo relativo) a un riesgo basal determinado. El segundo aspecto relacionado con el riesgo basal que necesita ser tomado en cuenta es el grado de variación que puede esperarse del impacto de la intervención. Por ejemplo, puede ser útil calcular el

número que es necesario tratar para el rango de riesgos basales observados en los grupos control de los estudios incluidos en una revisión.

9. 2. 4. Variación en los resultados de los estudios incluidos.

Además de estudiar la aplicabilidad de los resultados de una revisión, los revisores deberían analizar y valorar las variaciones importantes de los resultados en las condiciones en las que son aplicables. ¿Hay una variación previsible en los efectos relativos de la intervención? ¿Existen factores identificables que pueden modificar la respuesta o los efectos? Estos podrían incluir:

- Características del paciente, como la edad, sexo, marcadores bioquímicos.
- Características de la intervención, como la secuencia o la intensidad de la misma.
- Características de la enfermedad, como el estado de los receptores hormonales.

Estas características deben analizarse aunque no haya una heterogeneidad estadísticamente significativa. Debe hacerse estudiando si hay una interacción con el tratamiento y no mediante análisis de sub-grupos. Como ya se ha visto en la sección 8.7, es necesario interpretar cautelosamente las diferencias entre los sub-grupos, en particular las diferencias entre los estudios. Alguna variación casual entre sub-grupos es inevitable, por lo tanto, salvo que haya sólida evidencia de una interacción, debe asumirse que no existe.

9. 3. Otra información relevante.

Puede ser muy útil para los revisores valorar los resultados de una revisión en el contexto de otra información relevante, como es el caso de los datos epidemiológicos acerca de la magnitud y la distribución de un problema, la información acerca de la práctica clínica actual proporcionada por bases de datos institucionales o por encuestas sobre la práctica clínica, o bien la información acerca de los costes. Sin embargo, a menudo esto está fuera del alcance de las revisiones Cochrane y puede hacerse mejor en un entorno nacional o regional (por ejemplo, por las personas que desarrollan guías de práctica clínica o llevan a cabo evaluaciones de tecnología). Debe tenerse presente que la evidencia acerca de los efectos de la atención sanitaria es esencial para tomar decisiones bien informadas, pero no es suficiente. No puede y no debe esperarse que las revisiones Cochrane proporcionen toda la información que necesitan quienes toman decisiones. Por otro lado, los revisores pueden ayudar clarificando otras informaciones, que pueden variar ampliamente y que quizás resulten fundamentales en el momento de tomar una decisión.

9.4. Efectos adversos

La discusión y las conclusiones de una revisión deben mencionar la fuerza de la evidencia sobre los efectos adversos, incluyendo los estimativos de su gravedad y su frecuencia en diferentes situaciones. En particular, debe hacerse una evaluación crítica de la relación causal entre un efecto adverso y una intervención particular, teniendo en cuenta que una sub-evaluación y un subreporte de efectos adversos e inesperados son algo frecuentes. Los revisores pueden desear hacer comentarios acerca de cómo deben investigarse adicionalmente los efectos adversos en la sección de Implicaciones para la Investigación.

9. 5. Balance (*trade-off*).

Además de considerar el grado de evidencia que apoya cualquier conclusión, los revisores deben ser muy explícitos al establecer un orden de preferencias (valores asignados entre los diferentes resultados). Las intervenciones sanitarias provocan expectativas de obtención de beneficios pero tienen un precio en costes y riesgo de eventos adversos. Establecer conclusiones referentes a la utilidad práctica de una intervención implica sopesar, sea implícita o explícitamente, entre los beneficios esperados y los costes estimados y los posibles daños (Eddy 1990b). Está fuera de los objetivos de las revisiones Cochrane incorporar análisis económicos formales (aunque es bien posible que aquéllas se usen para tales análisis) (Mugford 1989, Mugford 1991). Sin embargo, los revisores deberían considerar todos los resultados potencialmente importantes de una intervención cuando saquen conclusiones, incluidos aquellos de los que es posible que no haya ningún dato fiable en los ensayos incluidos. También deberían ser cautelosos con cualquier asunción que hagan sobre el valor relativo de los beneficios, daños y costes de una intervención.

9. 6. Implicaciones.

Sin olvidar lo dicho acerca de las conclusiones inconsistentes, los grupos de revisión (y los usuarios de las revisiones Cochrane) quizá consideren útil clasificar las intervenciones en seis categorías mutuamente excluyentes. Esto ha sido realizado por el Grupo de Embarazo y Parto (Cochrane Pregnancy and Childbirth Group) (Enkin 1994) a partir de un trabajo previo de clasificación de las intervenciones en cuatro categorías que despertó gran interés y fue digno de encomio. Las tres primeras categorías de intervenciones, enumeradas abajo, son aquellas para las que hay suficiente evidencia disponible como para extraer conclusiones suficientemente firmes para la práctica. Las tres últimas son categorías para las que se necesita investigación adicional antes de que puedan extraerse sólidas conclusiones prácticas.

1. Prestaciones sanitarias de las que hay evidencia suficiente para establecer guías prácticas claras.

    A) Prestaciones sanitarias que mejoran los resultados.

    B) Prestaciones sanitarias que deben ser abandonadas considerando la evidencia disponible.

    C) Prestaciones sanitarias en que se produce un equilibrio entre beneficios y efectos adversos conocidos.

2. Prestaciones sanitarias para las cuales las evidencias disponibles son insuficientes si se desea proporcionar guías claras de práctica clínica, pero que deberían influir en las prioridades de la investigación.

    A) Prestaciones sanitarias que parecen prometedoras, pero que requieren una evaluación adicional.

    B) Prestaciones sanitarias en las que no se ha demostrado que tengan los efectos esperados, pero que merecen atención adicional.

    C) Prestaciones sanitarias con evidencia razonable de que no son eficaces para la finalidad para la cual se han aplicado.

9. 7. Errores comunes al establecer conclusiones.

Una equivocación común que ocurre, tanto al describir resultados como al sacar conclusiones -cuando la evidencia disponible no es concluyente- es confundir “no hay evidencia de efecto” con “evidencia de ningún efecto”. Los revisores deben ser cuidadosos de no caer en ello. Una segunda equivocación común es sacar conclusiones que van más allá de la evidencia que se analiza. Los revisores deben ser cautelosos y evitar este tipo de errores. Incluso afirmaciones generales del tipo "se necesita más investigación" también se deben evitar. Los revisores deben manifestar exactamente qué investigación se necesita y por qué. Las afirmaciones sobre cómo es posible mejorar la revisión con datos o recursos adicionales también pueden hacerse constar.

En resumen, no es posible proporcionar guías rígidas de cómo preparar una discusión y cómo llegar a conclusiones acerca de las implicaciones de una revisión para la práctica y para la futura investigación. Sin embargo, la consideración sistemática de los temas tratados en esta sección puede orientar la interpretación de los resultados de una revisión y ayudar a los revisores a evitar hacer inferencias incorrectas o engañosas.

9.8 Referencias

CTFPHE 1979. Canadian Task Force on the Periodic Health Examination. The periodic health examination. Can Med Assoc J 1979; 121:1193-254.

Cook 1992. Cook DJ, Guyatt GH, Laupacis A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Antithrombotic Therapy Consensus Conference. Chest 1992; 102:305S-311S.

Dans 1996. Dans AL, Dans LF. Users' guides for the applicability of the results of clinical trials, perspective of clinicians from developing countries. Manila: UP College of Medicine, 1996 (unpublished manuscript).

Eddy 1990b. Eddy DM: Anatomy of a decision. JAMA 1990; 263:4413.

Enkin 1994. Enkin M, Keirse MJNC, Renfrew MJ, Neilson JP. A Guide to Effective Care in Pregnancy and Childbirth. 2nd edition. Oxford: Oxford University Press, 1994.

Friedman 1985. Friedman LM, Furberg CD, DeMets DL. Fundamentals of Clinical Trials. 2nd edition. Littleton, MA: John Wright PSG Inc, 1985.

Guyatt 1994. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, II: how to use an article about therapy or prevention, B: what were the results and will they help me in caring for my patients. JAMA 1994; 271:59-63.

Guyatt 1995. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ, for the Evidence-Based Medicine Working Group. Users’ guide to the medical literature IX: A method for grading health care recommendations. JAMA 1995; 274:1800-4.

Gyorkos 1994. Gyorkos TW, Tannenbaum TN, Abrahamowicz M, Oxman AD, Scott EA, Millson ME, et al. An approach to the development of practice guidelines for community health interventions. Can J Public Health 1994; 85(Suppl 1):S8-S13.

Hill 1971. Hill AB. Principles of Medical Statistics. 9th edition. London: Lancet, 1971: 312-20.

Mugford 1989. Mugford M, Kingston J, Chalmers I. Reducing the incidence of infection after caesarean section: implications of prophylaxis with antibiotics for hospital resources. BMJ 1989; 299:1003-6.

Mugford 1991. Mugford M, Piercy J, Chalmers I. Cost implications of different approaches to the prevention of respiratory distress syndrome. Arch Dis Child 1991; 66:757-64.

US PSTF 1996. US Preventive Services Task Force. Guide to Clinical Preventive Services. 2nd edition. Baltimore: Williams & Wilkins, 1996; xxxix-lv.

10. MEJORA Y ACTUALIZACIÓN DE LAS REVISIONES.

Si las revisiones Cochrane aspiran a ser útiles para quienes desean tomar decisiones bien informadas sobre la atención sanitaria o la investigación, entonces aquellas deben ser fidedignas y transparentes. Tal y como se pone de manifiesto a lo largo de este Manual, la Colaboración utiliza métodos explícitos para producir revisiones y esta característica por sí sola las hará más útiles para los usuarios que la gran mayoría de las revisiones disponibles actualmente. No olvidemos que los libros de texto y los artículos de revisión con secciones de "Materiales" y "Métodos" siguen siendo raros.

Por encima de un cierto umbral mínimo de calidad, las revisiones que se han agregado a la Base de Datos Cochrane de Revisiones Sistemáticas-CDSR *(Cochrane Database of Systematic Reviews*) varían en el nivel de calidad metodológica al que hayan podido llegar los revisores respectivos. El "patrón-oro" (*gold-standard*) seguirá estando representada por las revisiones sistemáticas llevadas a cabo bajo el auspicio de los grupos colaboradores de investigación, y realizadas a partir de los datos individuales de todos los pacientes participantes en los ensayos incluidos en la revisión (ver sección 11). Tales revisiones requieren no sólo recursos sustanciales (incluido el tiempo), sino que también dependen del éxito de las negociaciones entre los investigadores. No deben subestimarse estos factores. Además, dado que "lo perfecto es enemigo de lo bueno", sería importante llevar a cabo investigación empírica para aprender más de lo que se sabe actualmente acerca de qué normas metodológicas son esenciales, y cuáles son deseables, para evitar los sesgos.

Los mecanismos para mantener y elevar los estándares de las revisiones Cochrane incluyen:

- Atraer participantes con dedicación y evitar conflictos de intereses.
- Participación de los consumidores.
- Garantizar el acceso a los estudios.
- Mejorar el acceso a los datos no publicados.
- Establecer y desarrollar normas y pautas de actuación.
- Uso de métodos rigurosos de revisión.
- *Software* y apoyo informático.
- Formación de los revisores.
- Revisión por pares (*peer-review*) abierta y continua.
- Actualización de las revisiones.

10. 1. Atraer participantes con dedicación y evitar conflictos de interés

La calidad de la Colaboración Cochrane, desde cualquier punto de vista, será un reflejo de las características de los individuos que la componen. Actualmente, estas personas se han seleccionado por sí mismas durante las fases más tempranas del desarrollo de la Colaboración. El colectivo de personas que tienen experiencia en preparar revisiones sistemáticas sobre los efectos de la atención sanitaria sigue siendo pequeño, y entre los que tienen esta experiencia, sólo algunos desearán asumir el compromiso sustancial que se espera de aquel que entra a formar parte de la Colaboración. En otras palabras, los obstáculos que hallan los que están preguntándose si deben tomar parte o no en la Colaboración son en sí mismos una forma de *screening* (cribaje, tamizaje).

A fin de asegurar la integridad real y percibida de las revisiones Cochrane, la Colaboración ha adoptado un Código de Conducta para evitar potenciales conflictos

de intereses financieros. Los revisores deben evitar conflictos de intereses financieros y revelar aquellos que no puedan evitarse (ver sección 2. 2). Asimismo, todos los revisores deben firmar las declaraciones de responsabilidad y de conflicto de intereses. El equipo editorial de cada Grupo Colaborador de Revisión (CRG) también debe revelar cualquier potencial conflicto de intereses.

10. 2. Participación de los Consumidores.

Los consumidores de atención sanitaria y otros usuarios de las revisiones Cochrane deben participar en desarrollar revisiones para asegurar que éstas:

- Se refieran a problemas que atañen a la gente.
- Consideren resultados que son importantes para los afectados.
- Sean accesibles a las personas que toman decisiones.
- Reflejen adecuadamente la variabilidad en los valores y condiciones de las personas.

Se sabe relativamente poco acerca de la eficacia de las diversas maneras de incluir a los consumidores en el proceso de revisión o, más en general, en la investigación sanitaria, desde cómo se formulan las preguntas a cómo se usan sus resultados en las decisiones. Desde un principio la Colaboración está comprometida en la participación de los consumidores, la cual se basa en nuestros valores y lógica (Bastian 1998). Los investigadores y quienes financian la investigación no han sido capaces de asegurar que la investigación sanitaria satisfaga adecuadamente las necesidades de aquellos que, en último término, son los afectados. Debido a los valores o intereses en conflicto, es improbable que esta situación vaya a mejorar sustancialmente si no se dispone de mecanismos apropiados que incluyan a los consumidores en las decisiones referidas a la investigación. Para asegurar la eficacia de su participación, deben usarse la creatividad y un enfoque crítico para desarrollar y evaluar los mecanismos que se establezcan para ello. Esto está llevándose a cabo de muchas maneras por los CRG, a través de las actividades de la Red de Consumidores Cochrane (*Consumers' Network*) y por otras entidades dentro de la Colaboración. Esta experiencia práctica y sus evaluaciones formales proporcionarán una base para establecer normas sobre cómo asegurar que la participación de los consumidores contribuya eficazmente a asegurar la calidad y la accesibilidad de las revisiones Cochrane.

Los consumidores están participando en el desarrollo de protocolos y revisiones de la siguiente manera:

- Ayudando a determinar temas y aspectos para las revisiones
- Como co-revisores
- Como parte de las consultas al consumidor durante el desarrollo del protocolo y de la revisión (mediante cuestionarios, diálogo directo o entrevistas, en grupos focales y en grupos de discusión por correo electrónico o teleconferencias).
- Como árbitros durante el proceso editorial.

Cuando se consulta a los consumidores (o a otras personas) durante el desarrollo de un protocolo o una revisión, deben agradecerse sus contribuciones en la sección de Agradecimientos del protocolo o de la revisión. También puede ser apropiada su inclusión formal en la lista de revisores citados, como se hace con otras personas que hicieron contribuciones.

También puede discutirse dentro de la revisión la literatura elaborada por consumidores, encuestas y estudios que exploran los puntos de vista de los

consumidores, para garantizar que se abordaron aspectos que son importantes para los consumidores.

Muchos de estos elementos también se aplican a otros usuarios de las revisiones Cochrane. La Colaboración está igualmente comprometida con la participación de los usuarios e invita a los revisores a que busquen e incorporen los puntos de vista de otros usuarios además de los consumidores en el desarrollo de protocolos y revisiones.

10. 3. Asegurar el acceso a los estudios.

Debido al desorden que reina en la bibliografía médica, se requiere una dosis considerable de esfuerzo para localizar la investigación que aborda los temas planteados en una revisión (ver sección 5). La Colaboración está ayudando a asegurar que los estudios válidos y pertinentes, sean ubicados por los revisores e incluidos en sus revisiones, mediante las siguientes acciones:

- Realización de búsquedas manuales de la bibliografía sanitaria en todo el mundo para identificar los ensayos.
- Apoyo al desarrollo y el mantenimiento de los registros de ensayos que realizan los CRG.
- Brindar formación y apoyo a aquellas personas que emprenden actividades de búsqueda.
- Desarrollo del Registro de Ensayos Controlados Cochrane-CCTR (*Cochrane Controlled Trials Register*) para facilitar la transferencia de ensayos entre los CRG y otras entidades Cochrane y para facilitar el acceso a los ensayos en otros registros que hacen su aportación al CCTR.
- Colaboración con la Biblioteca Nacional de Medicina de los Estados Unidos (US National Library of Medicine) para mejorar la codificación de los ensayos en MEDLINE y desarrollar una base de datos auxiliar de ensayos no incluidos en MEDLINE.
- Desarrollo y evaluación de estrategias para mejorar la codificación y clasificación de los ensayos.

Estos trabajos incluyen a un gran número de personas dedicadas a una gran variedad de actividades a través de los CRG, de los Centros Cochrane, los Ámbitos (*Networks*) y los Grupos de Trabajo sobre Metodología (*Methods Working Groups*). El Grupo de Desarrollo de los Registros de Ensayos (*Trials Registers Development Group*), el Centro Cochrane de New England (*Providence Office*) y el Grupo Métodos sobre recuperación de información (*Information and retrieval Methods Group*) tienen responsabilidades clave en la coordinación y guía de estas actividades.

10. 4. Mejora del acceso a los datos no publicados.

Se necesita tener un mejor acceso a los datos no publicados para superar los problemas que acarrea la falta de información de los informes publicados y para protegerse contra el sesgo de publicación. Además de los esfuerzos emprendidos por cada CRG para ayudar a garantizar el acceso a datos no publicados válidos dentro de su área, toda la Colaboración está trabajando para desarrollar alianzas estratégicas con la industria farmacéutica y otras, y está promoviendo activamente unas normas éticas que establecen lo inaceptable que resulta retener los datos no publicados.

10. 5. Establecer y desarrollar normas y guías.

El Manual es la expresión más tangible del desarrollo de normas y guías de la Colaboración Cochrane. Dado que ya se acumula una cierta experiencia entre las personas que tratan de aplicarlas, y a la luz de los resultados válidos de la investigación, el Manual ya ha sido y seguirá modificándose. Además de basarse en la experiencia de los CRG, múltiples fuentes contribuyen al desarrollo del Manual; entre otras: estudios metodológicos empíricos, otros artículos metodológicos pertenecientes a la base de datos de revisiones Cochrane sobre Metodología (*Cochrane Review Methodology Database*) y el asesoramiento de los Grupos de Trabajo sobre Metodología correspondientes.

Más allá de lo que puedan encontrar en el Manual, cada CRG debe tomar decisiones acerca de las normas y pautas adecuadas a la naturaleza de los problemas de atención sanitaria de que se ocupen. Los métodos estándar usados por un CRG se publican en el módulo del Grupo en la Cochrane Library.

10. 6. Utilización de métodos rigurosos de revisión.

No es ni factible ni aconsejable dictar las decisiones que debería tomar un revisor. Estas variarán de revisión a revisión según el tema, la naturaleza de los ensayos disponibles y los recursos a su alcance. Sin embargo, en general la validez de las revisiones Cochrane está asegurada por:

- Búsqueda tan completa como sea posible de los estudios que cumplen criterios de inclusión en una revisión, basándose en la mayor medida posible en los esfuerzos centralizados para ayudar en esta tarea y asegurando meticulosidad y eficiencia para la identificación de los ensayos clínicos aleatorios.
- Uso de criterios explícitos en la selección de ensayos para su inclusión en una revisión y para la evaluación de su calidad.
- Aplicación de estos criterios por más de un revisor cuando sea conveniente y factible a fin de asegurar la posibilidad de reproducir los juicios que se emitan.
- Esfuerzos continuados para recopilar la información que falte y que es posible que contribuya de forma importante a una revisión, en tanto y en cuanto la disponibilidad de datos y de recursos lo permita.
- Recopilación de datos de pacientes individuales de los investigadores cuando sea conveniente y factible, también según la disponibilidad de recursos y datos.
- Uso de técnicas estadísticas apropiadas, según corresponda, para sintetizar los resultados.
- Uso de análisis de sensibilidad para poner a prueba la solidez de los resultados de una revisión en relación con cualquier juicio o asunción.
- Uso prudente de los análisis de subgrupos y evitar la sobreinterpretación de cualquier análisis de subgrupos que se realice.
- Extraer conclusiones con cuidado, incluidas las implicaciones para la práctica e investigación futura, basándose en una interpretación precavida de los resultados, considerando las limitaciones de la revisión y la variabilidad en los valores y condiciones de los que toman las decisiones.
- Información completa de los materiales y métodos usados al emprender la revisión.

Así como es posible actualizar las revisiones Cochrane atendiendo a nuevas evidencias disponibles, también es posible mejorar los métodos. Además, dado que los métodos se describen explícitamente en las revisiones Cochrane, los usuarios pueden juzgar por sí mismos la validez de los resultados de una revisión.

10. 7. *Software* y apoyo informático.

El programa *Review Manager* (RevMan) está diseñado para ayudar a los revisores a construir las revisiones según un formato estructurado descrito en la sección 2. Este *software* seguirá desarrollándose e incorporará normas y guías para las revisiones Cochrane, mejores métodos analíticos, una ayuda "en línea" (*on-line*) y mecanismos de control de errores, según éstos vayan evolucionando. Una manera mediante la cual el *software* contribuye a asegurar la validez de las revisiones Cochrane es facilitando *a priori* el registro de los protocolos de las revisiones previstas para el futuro, según se describe en la sección 3.

El desarrollo del RevMan, del *Module Manager* (ModMan) y de la *Parent Database* (base de datos matriz) (conjunto de programas de computación usados para preparar y compilar la CDSR) está dirigido por el Grupo de Desarrollo de *Software* (*Software Development Group*) con la orientación de diversos grupos de consulta para cada uno de los tres programas. Actualmente está explorándose el desarrollo de *software* especializado para facilitar el manejo de los ensayos por los CRG (TrakMan). El asesoramiento, apoyo y entrenamiento en lo referente al uso de la informática, Internet y de los métodos que satisfagan las necesidades de información de la Colaboración son proporcionados por el Grupo de Trabajo de Métodos Informáticos (*Informatics Methods Working Group*). Los Centros Cochrane de Canadá, de Australia y Nórdico han asumido responsabilidades específicas destinadas a satisfacer estas necesidades. Se puede encontrar información acerca de las actividades de cada una de estas entidades en los módulos respectivos que figuran en la CDSR.

10. 8. Formación.

Es importante asegurar que las personas que participan en la Colaboración tengan los conocimientos y las capacidades que necesitan para hacer un buen trabajo. El adiestramiento puede ser necesario para los revisores, los editores, los coordinadores de los CRG, para quienes realizan búsquedas manuales, los formadores y los usuarios de las revisiones Cochrane. Nos centramos aquí en las necesidades de aprendizaje de los revisores y los editores para ayudarles a preparar y a actualizar revisiones de alta calidad.

10. 8. 1. Formación para revisores.

Si bien algunos revisores que se unen a un Grupo Colaborador de Revisión tienen conocimientos y experiencia para realizar una revisión sistemática, muchos de ellos no. Los Centros Cochrane son responsables de la preparación de los materiales de aprendizaje y de la organización de seminarios de formación para los CRG. Cada CRG se encarga de asegurar que los miembros del grupo tengan el adiestramiento y el apoyo metodológico adecuados. Continuamente aparecen materiales de estudio y oportunidades de formación, que evolucionarán como reflejo de las necesidades de la Colaboración, sus normas y pautas.

10. 8. 2. Formación para los editores.

Los editores de los CRG necesitan tener experiencia relacionada con el área que cubre el grupo de revisión, habilidad y experiencia en avaluar críticamente los estudios sobre los efectos de la atención sanitaria y capacidad para editar el material científico que se publique. Un editor de un CRG debería:

- Identificar y definir los temas que se van a revisar.
- Identificar a los potenciales revisores que sean capaces de preparar una revisión sistemática y tenerlos informados.

- Apoyar y ayudar a los revisores a superar las dificultades que encuentren en la preparación de sus revisiones.
- Controlar y asegurar que las revisiones tengan sentido científico y encaren temas relevantes, utilizando revisores independientes cuando parezca apropiado.
- Controlar y asegurar que las revisiones sean realizadas en el formato tipo que utiliza la Colaboración Cochrane.
- Organizar la actualización de las revisiones existentes, mediante la recuperación de estudios recién registrados dentro del área del CRG, enviándoselos a los revisores y ayudándoles a actualizar sus revisiones.

Algunas de estas tareas son comunes a muchos tipos de trabajo editorial médico, otras son específicas al trabajo global en la Colaboración Cochrane y unas pocas a áreas específicas o a los CRG. La formación de los editores está por tanto constituida por una formación editorial general, formación dentro de la Colaboración Cochrane y formación respecto al trabajo dentro de los CRG.

Probablemente, el adiestramiento general sea proporcionado más eficientemente en cursos, talleres y seminarios dirigidos a editores científicos -por ejemplo, por la Asociación Europea de Editores de Ciencia (*European Association of Science Editors-EASE*) o el Consejo Norteamericano de Editores de Biología (*North American Council of Biology Editors-CBE*). El Centro Cochrane del Reino Unido es un miembro de la EASE; quizá sea útil para otros Centros Cochrane acogerse a la EASE o al CBE.

El entrenamiento dentro de la Colaboración Cochrane toma la forma de talleres y seminarios organizados por los Centros Cochrane, con participación de editores experimentados de los CRG ya establecidos y editores principiantes. Algunos de estos talleres supondrán compartir, en general, una experiencia editorial, otros quizá se centren en aspectos especializados y concretos de las revisiones (como puede ser el uso correcto del metanálisis, cómo planificar el trabajo de un Grupo Colaborador de Revisión o cómo ayudar y apoyar a los revisores).

Dentro de cada CRG, el o los editores veteranos deben asumir la responsabilidad de potenciar el desarrollo sistemático de las aptitudes editoriales de los otros editores. Dado que esta actividad será nueva para la mayoría de los editores veteranos, quizá sea útil para la Colaboración Cochrane organizar talleres introductorios.

Cuando un nuevo editor o editora se incorpora a un grupo, los editores experimentados deben evaluar su experiencia y aptitudes. El nuevo miembro debe empezar editando una o varias revisiones junto con uno de los miembros más experimentados, el cual puede actuar como supervisor o tutor. Juntos pueden hacer un plan esquemático para la formación del nuevo editor, revisándolo a medida que el proceso de formación avanza. Quizá sea útil para la coordinación de la oficina editorial de un CRG mantener un registro formal o informal de la experiencia de trabajo y el adiestramiento de cada editor.

10. 9. Revisión por pares (*peer-review*) y Sistema de Manejo de Críticas.

Es importante tener mecanismos eficientes para la crítica de las revisiones realizadas por los colaboradores de la Colaboración Cochrane y para modificar las revisiones atendiendo a las críticas que sean válidas. El desarrollo de estos mecanismos se facilita si la potencialidad de la publicación electrónica se explota imaginativamente. Tradicionalmente, las oportunidades de criticar las revisiones antes de que se publiquen impresas en papel quedan restringidas por el número y competencia de los revisores que seleccionan los editores. Después de que se haya publicado una revisión, las oportunidades de publicar las críticas quedan limitadas generalmente a

las pocas cartas que los editores pueden aceptar para su publicación o a las reseñas, que a menudo son excesivamente breves e inespecíficas. También resulta desalentador que no haya ninguna manera directa mediante la cual los autores de las revisiones impresas puedan modificar sus revisiones tras considerar las críticas válidas que les llegan.

La Colaboración Cochrane intenta crear un sistema iterativo mediante el cual las sucesivas versiones de cada revisión reflejarán no solo la aparición de nuevos datos, sino también las críticas válidas -solicitadas o espontáneas- de cualquier fuente. Este proceso se ve facilitado por los detalles de contacto, tanto de los revisores como de las sedes editoriales, que aparecen en cada revisión. Las versiones sucesivas de una determinada revisión, junto con cualquier crítica realizada, se archivarán electrónicamente.

10. 9. 1. Revisión.

Se requiere que cada CRG publique una declaración que describa su política de revisión por pares antes de la publicación, en la sección “Proceso Editorial” de su módulo en la Cochrane Library.

Los principales aspectos que hay que tener en cuenta cuando se está considerando el registro del título de una revisión son si existe algún tipo de solapamiento o duplicación potencial de esfuerzos con otro revisor, dentro o fuera del CRG; si se han redactado claramente los objetivos y si estos incluyen todos los componentes de una pregunta bien formulada; y si la revisión tiene posibilidades de ser llevada a cabo. En esta etapa, la valoración generalmente puede realizarse rápidamente por el equipo editorial del CRG.

La revisión de los protocolos puede ser más lenta que la evaluación de la revisión completa. Ésta se realiza para asegurar que la información contenida en los antecedentes sea racional y esté presentada con claridad, y que se hayan planificado métodos apropiados para identificar, obtener y sintetizar los datos. En esta etapa, la revisión por pares es particularmente importante para prevenir errores metodológicos que quizás no sean remediables fácilmente en etapas posteriores. La revisión por pares de la revisión completa incluye una segunda crítica de los métodos de la revisión así como una crítica de los resultados obtenidos, de la presentación de resultados, del debate y de las conclusiones.

Antes de su publicación, todas las revisiones deben ser evaluadas por lo menos por dos personas distintas a los editores del CRG. Los editores del CRG asignan un revisor árbitro (o de contacto) para cada revisión. Si informan al Centro Cochrane de San Francisco (San Francisco Cochrane Center) (sfcc@sirius.com) acerca de su elección, el Centro puede entrenar y apoyar a esta persona. Se recomienda que dichos árbitros tengan: 1) experiencia metodológica, 2) experiencia con relación al área de contenido, y/o 3) sea un consumidor potencial de la revisión. Los dos árbitros deben seleccionarse contando con que tengan diferentes puntos de vista. Los árbitros deben ser personas que no tengan conflictos de intereses directos de tipo personal o financiero con respecto al tema que se está abordando. Debe pedirse a los árbitros que envíen comentarios corteses y constructivos acerca de la Revisión, que identifiquen las debilidades o los errores graves de ésta, así como la manera de mejorarlos. También debe solicitarse que remitan estos comentarios al editor árbitro al cabo de un mes como máximo.

Debe estimularse entre las personas que hacen la revisión por pares el uso de métodos estandarizados y listas que sean explícitas con el fin de garantizar que la

información sea comprensible y que se limitan los sesgos. Las áreas específicas que se abordan en cada fase de la revisión por pares son variables. Las diferencias entre las críticas de los revisores externos deben dilucidarse y conciliarse siempre que sea posible. Esto puede realizarse mediante el arbitraje por parte de los editores del CRG o con otros árbitros independientes.

El editor árbitro debe controlar el cronograma en que se pretende devolver los comentarios, el grado de calidad de los comentarios y, en caso de ser necesario, nombrar un revisor de reserva. Envia los comentarios de los árbitros junto con sus propios comentarios (en caso de que existan) a los autores de la revisión o al coordinador del CRG para que se distribuyan entre los autores de la revisión y, cuando sea apropiado, entre los demás editores. El editor árbitro, junto con el equipo editorial, aprueba la versión final de la revisión antes de que sea publicada en la Cochrane Library.

El editor árbitro debe mantener un registro de todos los materiales recibidos y enviados durante el proceso de arbitraje. Se está desarrollando un sistema electrónico de arbitraje para mantener el registro electrónico de estos intercambios. El Centro Cochrane de San Francisco (San Francisco Cochrane Center) solicitará copias de los registros y los estudiará periódicamente con el fin de mejorar el proceso de arbitraje.

10. 9. 2. Lista de Verificación (*checklist*) para la revisión por pares.

Preparar una revisión supone emitir juicios en cada etapa del proceso de revisión. Pueden cometerse errores tanto sistemáticos como aleatorios. Existen varias listas de verificación (*Checklist*) que permiten a los revisores usarlas como guías en la detección de errores importantes en el proceso de revisión. Algunos puntos que deben tenerse en cuenta se presentan más adelante. Han sido recogidos a partir de múltiples referencias (Jackson 1980, Cooper 1982, Light 1984a, L'Abbe 1987, Mulrow 1987, Sacks 1987, Oxman 1988, Oxman 1994a, Oxman 1994b, Cook 1995).

<u>Formulación de la pregunta.</u>

- Las preguntas de la revisión, ¿están bien formuladas y contienen los componentes clave?
- ¿Se documentaron y justificaron adecuadamente los cambios en el protocolo?

<u>Identificación de estudios.</u>

- ¿Hay una búsqueda minuciosa de datos pertinentes usando fuentes apropiadas?
- ¿Las estrategias de búsqueda son adecuadas para la pregunta que se ha planteado?

<u>Selección de estudios.</u>

- ¿Se aplican criterios de inclusión y de exclusión apropiados para seleccionar los estudios?
- ¿Se aplican los criterios de selección de manera que limiten los sesgos?

<u>Evaluación de los estudios.</u>

- ¿Se valora de manera fiable la validez de los estudios individuales?
- Los parámetros importantes que pueden afectar a los resultados del estudio (por ejemplo, ámbito, población de estudio, diseño del estudio), ¿se abordan de modo sistemático?

Obtención de datos.

- ¿Hay una mínima cantidad de datos no disponibles en los resultados y en otras variables que se consideren clave para la interpretación de los resultados?

Síntesis de datos.

- ¿Se tomaron decisiones razonables con respecto a la manera de combinar los datos?
- ¿Se han considerado en la síntesis factores importantes como el diseño de los estudios?
- ¿Son sensibles los resultados a los cambios relacionados con la manera con que se hizo el análisis?
- ¿Se informa de la precisión de los resultados?

Discusión.

- ¿Se mencionan las limitaciones de los estudios y del proceso de revisión?
- ¿Se integran los resultados de la revisión en el contexto de otra evidencia indirecta?

Conclusiones de los revisores.

- Las conclusiones, ¿se apoyan en el contenido de la revisión?
- ¿Se abordan explicaciones alternativas y verosímiles de los efectos observados?
- ¿Hay interpretaciones adecuadas acerca de la evidencia que no sea concluyente (por ejemplo, que no hay evidencia del efecto y/o evidencia de que una estrategia particular no funcione (evidencia de que no hay efecto)?
- ¿Se identifican las consideraciones que resultan importantes para las personas que deben adoptar decisiones, incluidos los valores y factores contextuales que pueden influir en sus decisiones?

10. 10. Actualización de las revisiones.

Cuando se registra una revisión ante la Colaboración Cochrane, los revisores se comprometen a mantenerla actualizada. Mantener actualizada una revisión implica la repetición, a intervalos periódicos, de las etapas comprendidos en la revisión original. Algunas de las etapas requerirán un esfuerzo mínimo (por ejemplo, analizar la pregunta de investigación para asegurar que todavía está vigente) mientras que otros quizá requieran una inversión considerable de tiempo y esfuerzo.

El aspecto logísticamente más exigente para mantener una revisión actualizada es la identificación de los nuevos estudios. Para los CRG que se organizan bien y están suficientemente financiados, la identificación periódica de nuevos estudios elegibles es una función del equipo editorial (habitualmente el coordinador del CRG o el coordinador de Búsquedas de Ensayos). En otros casos, los revisores y editores deben elaborar mecanismos de colaboración para identificar periódicamente nuevos estudios. Como mínimo, las estrategias para identificar los nuevos estudios deben incluir periódicamente la consulta del registro del CRG, CENTRAL/CCTR y MEDLINE. La Colaboración Cochrane ha implantado el sistema de registro de críticas (*Criticism Management System*, ver más abajo), que continúa siendo desarrollado y mediante el cual los usuarios de las revisiones Cochrane hacen comentarios y críticas a las revisiones, tal como se discute más adelante. Es probable que sea una fuente adicional de estudios para ser considerados en una revisión.

Los formularios para la recogida de datos originales deben usarse para resumir las evidencias de nuevas investigaciones. Si la evidencia proveniente de nuevas investigaciones trata variables importantes que no estaban incluidas en el formulario de recopilación original, habría que modificar estos formularios. Por ejemplo, si originalmente los revisores sólo hubieran incluido los resultados de mortalidad y morbilidad en los ensayos que estudiasen el tratamiento de cáncer avanzado y los estudios recientes proporcionasen rutinariamente resultados sobre calidad de vida, el formulario de recogida de datos podría modificarse. En tales casos, los revisores quizá necesiten volver a verificar si alguno de los estudios que identificaron antes tenía esa información y fue pasada por alto.

A veces los revisores toman la decisión de incluir una nueva estrategia de análisis en la actualización de su revisión; por ejemplo, utilizando métodos estadísticos que anteriormente no estaban disponibles en el RevMan. En general, las nuevas estrategias de análisis supondrán cambios significativos que merecen una crítica editorial a través del proceso editorial establecido por el CRG.

La frecuencia con que una revisión necesita ser actualizada variará según la aparición de nuevas evidencias fruto de trabajos de investigación válidos. Los revisores deben trabajar con el comité editorial de su CRG para establecer las pautas a seguir cuando una nueva evidencia proveniente de la investigación científica es lo bastante significativa como para requerir una actualización o una modificación sustancial de la revisión. Las fechas de tales modificaciones deben registrarse en la sección de "Qué es Nuevo" de la revisión. Es política de la Colaboración que los revisores actualicen anualmente la revisión o que añadan un comentario para explicar por qué esto se hace con menos frecuencia. Aunque no se encuentre ninguna nueva evidencia significativa en la actualización anual y no se consigne ninguna modificación mayor, esta información debe utilizarse para actualizar la revisión, añadiendo la fecha de la última búsqueda de evidencia de la revisión.

10. 11. Respondiendo a las críticas.

El formato electrónico de la CDSR ofrece una oportunidad única para responder e incorporar las críticas de los usuarios de las revisiones Cochrane. Esto aumentará enormemente la calidad de las revisiones y permitirá a los usuarios participar dentro del proceso de revisión.

El lector debe utilizar el botón de "Comentarios/Críticas" para hacer críticas constructivas y comentarios corteses. Éstos se envían automáticamente al Editor de Críticas del CRG correspondiente. Como una forma de evitar redundancias, el usuario debe leer las críticas que ya han sido recibidas antes de enviar las suyas. Esto puede hacerse al visitar la página de Internet de "Comentarios y Críticas" http://www.cochrane.de/cc/cochrane/currcrit.htm.

Cuando se reciban los comentarios, el Editor de Críticas debe resumirlos y enviar una copia a los revisores y al Centro Cochrane de San Francisco de modo que puedan aparecer en la página *web* de "Comentarios y Críticas". Los revisores son responsables de responder puntualmente todas las críticas. Deben proporcionar una respuesta escrita utilizando la sección de críticas de RevMan y actualizar su revisión en caso de ser apropiado. Se está desarrollando un programa informático para ayudar a los Editores de Críticas a que coordinen las respuestas de los revisores a los comentarios y las críticas.

10.12 Referencias

Bastian 1998. Bastian H. Speaking up for ourselves: the evolution of consumer advocacy in health care. International Journal of Technology Assessment in Health Care 1998;14:3-23.

Cook 1995. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-Analysis. J Clin Epidemiol 1995; 48:167-71.

Cooper 1982. Cooper HM. Scientific guidelines for conducting integrative research reviews. Rev Educ Res 1982; 52:291-302.

Jackson 1980. Jackson GB. Methods for integrative reviews. Rev Educ Res 1980; 50:438-60.

L'Abbe 1987. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Ann Intern Med 1987; 107:224-33.

Light 1984. Light RJ, Pillemer DB. Summing Up: The Science of Reviewing Research. Cambridge: Harvard University Press, 1984.

Mulrow 1987. Mulrow CD. The medical review article: state of the science. Ann Intern Med 1987; 106:485-8.

Oxman 1988. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. Can Med Assoc J 1988; 138:697-703.

Oxman 1994a. Oxman AD, Cook DJ, Guyatt GH. Users' guide to the medical literature, VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA 1994; 272:1367-71.

Oxman 1994b. Oxman AD. Checklists for review articles. BMJ 1994; 309:648-51.

Sacks 1987. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med 1987; 316:450-5.

11. REVISIONES QUE USAN DATOS DE PACIENTES INDIVIDUALES (DPI).

11. 1. Justificación.

Si una revisión sistemática va a incluir un metanálisis en el cual los resultados de estudios individuales se combinarán en una síntesis estadística, los datos podrían recogerse de diversas maneras. Puede optarse por la extracción de datos de los informes publicados, la recopilación de datos agregados proporcionados por los investigadores responsables u obtener de estos investigadores los datos de pacientes individuales DPI (*individual patient data*). Esto último se ha usado en revisiones cooperativas a gran escala en las cuales se combinan los datos de todos los ensayos aleatorios de un área específica de enfermedades (EBCTCG 1992) y también en revisiones más selectivas en las que se recogen y combinan los datos de un número relativamente pequeño de ensayos para evaluar una intervención sanitaria específica (Jeng 1995). Las revisiones sistemáticas basadas en DPI se han descrito como la referencia con la que deben compararse todas las revisiones (Chalmers 1993). Aunque pueden necesitar más tiempo, recursos y experiencia que otras formas de revisión, el proceso trae consigo algunas ventajas. Los revisores deben considerar la importancia de estas ventajas para su revisión sistemática en particular, al decidir si emprenden un proyecto de esta naturaleza. En la *Cochrane Database of Systematic Reviews* hay ejemplos de revisiones con DPI, incluyendo algunas que fueron originalmente publicadas en artículos de revistas.

11. 2. Grupo de Métodos sobre revisiones con datos de pacientes individuales.

Para tratar de ayudar a tomar esta decisión y con la logística propia de este tipo de grupos, se ha establecido en la Colaboración Cochrane un Grupo de Métodos (*Cochrane Collaboration Methods Group*) para ofrecer orientación a quienes deseen desarrollar un metanálisis con DPI. Este grupo (por Lesley Stewart y Mike Clarke) se formó después de un seminario patrocinado por el Centro Cochrane del Reino Unido en abril de 1994 en el cual se reunieron por primera vez representantes de los grupos de investigación interesados en tales proyectos. Esto permitió la discusión de algunos aspectos como el uso y desarrollo de protocolos, los métodos de validación de datos y los recursos necesarios. Un informe detallado de este seminario se publicó en la revista Statisics in Medicine en octubre de 1995 (Stewart 1995). La versión completa de este reporte se incluye en este Manual (Apéndice 11a) con autorización de la revista.

11. 3. ¿Qué es y qué no es un metanálisis con DPI?.

Del mismo modo que en cualquier revisión sistemática, el principio fundamental de una revisión que utilice datos de pacientes individuales (DPI) es que incorpora, en la medida de lo posible, toda la evidencia relevante y pertinente. Esto significa que el proceso de identificación de ensayos debe ser lo más minucioso posible y que el esfuerzo de recopilar sus datos debe ser igual de minucioso. El objetivo principal es que todos los participantes aleatorizados y ninguno de los no aleatorizados, de cualquier ensayo válido, sean incluidos y que se analicen usando el principio de la "intención de tratar" (*intention to treat*). De esta forma, los sesgos sistemáticos y los efectos del azar se reducirán al mínimo. Con este fin, la recopilación de datos debe mantenerse sencilla y directa, recogiendo la mínima cantidad de datos para los análisis necesarios. Debe hacerse lo más fácil posible a los investigadores el suministro de sus datos, ya que esto aumentará la probabilidad de que se obtengan los datos de todos los estudios válidos. Además, los investigadores deben saber que

cualquier dato suministrado para la revisión se mantendrá confidencialmente, no se usará para otra finalidad sin su permiso, y que los informes de la revisión se publicarán en nombre de los investigadores colaboradores en lugar de los coordinadores centrales.

La diferencia fundamental entre un metanálisis con DPI y un metanálisis basado en datos agregados (ya sean extraídos de informes publicados o proporcionados directamente por los investigadores) radica en que, en el primero, los resultados combinados de los estudios provienen de un reanálisis central de los datos brutos (*raw data*) de cada estudio. En primer lugar se buscan los datos necesarios de cada ensayo, y después del procesamiento central, se discute cualquier incongruencia o problema y se trata de resolver mediante la comunicación con los investigadores responsables. Luego, los datos finales de cada ensayo se analizan por separado para obtener estadísticos de resumen, los cuales se combinan para dar un estimador global del efecto del tratamiento. De esta forma, los participantes se comparan directamente sólo con otros del mismo estudio, y todo el conjunto de datos no se fusiona como si proviniera de un único estudio.

11. 4. ¿Qué ventajas ofrece un metanálisis con DPI?

Una revisión sistemática que dependa exclusivamente de los datos de los estudios publicados está expuesta a varios problemas. El más obvio de estos es que los estudios inéditos no estarán incluidos, pero también puede suceder que los datos publicados sean inadecuados. Por ejemplo, es posible que haya información insuficiente sobre los tipos de pacientes o resultados de interés en la revisión; los datos están "detenidos en el tiempo", cuando los resultados importantes quizá provengan de un seguimiento más largo o de un estudio más detallado; y el principio de la intención de tratar quizá no se haya seguido (a veces puede que no quede claro en el informe publicado). La recopilación de los datos agregados o de pacientes individuales directamente de los investigadores resolverá algunos de estos problemas: los ensayos inéditos pueden ser incluidos, pueden solicitarse los datos actualizados sobre tipos específicos de participantes y resultados, y puede aclararse si los datos se basan en asignaciones aleatorias (si los estudios son ensayos aleatorios).

La utilización de DPI en lugar de datos agregados conlleva ventajas adicionales. Éstas incluyen la posibilidad de realizar análisis de la supervivencia y otros análisis del tipo “tiempo hasta el evento” (*time to event analyses*); emprender análisis usando subgrupos comúnmente definidos para probar y generar hipótesis; asegurar la calidad de la aleatorización y los datos de seguimiento usados en el metanálisis, mediante la validación de los datos y la corrección iterativa de errores realizada a través de la comunicación directa con los investigadores; y actualizar la información de seguimiento a través de los sistemas de registro de pacientes (como los registros de mortalidad) cuando están disponibles. Además, quizá sea más fácil para un investigador enviar datos de pacientes individuales en lugar de datos resumidos del estudio, particularmente si no tienen suficiente apoyo estadístico o para gestionar los datos en la preparación de las tablas necesarias. También será más fácil si se trata de suministrar una cantidad pequeña de información extra. Por ejemplo, si el seguimiento adicional de algunos participantes está disponible, sencillamente el investigador puede enviar esta información en vez de preparar tablas nuevas.

Además, como los metanálisis con DPI requieren de la colaboración de los investigadores, pueden tener otros beneficios, algunos de los cuales también pueden encontrarse si se solicitan datos agregados a los investigadores. Dentro de estos, se incluyen la identificación y la comprensión más completa de los estudios; más facilidades para obtener información de los datos que falten (missing data); una

interpretación más equilibrada de los resultados de la revisión; mayores garantías y difusión de estos resultados; un consenso más amplio en las implicaciones para la práctica y la investigación futuras; y la posible colaboración en estas investigaciones.

11. 5. ¿Dónde está la evidencia?

Uno de los objetivos del Grupo de Métodos respecto a los metanálisis basados en DPI fue establecer y promover un programa de investigación para explorar este enfoque de cara a las revisiones sistemáticas. Ya existe alguna evidencia empírica, aunque limitada, de algunas de las ventajas de las revisiones con DPI sobre otros tipos de revisión. De manera característica, estas evidencias han incluido la comparación de los resultados de un metanálisis con DPI con los de un metanálisis basado en material publicado y han demostrado la importancia del primero para ayudar a controlar el sesgo de publicación, para asegurar la aplicación del principio de la intención de tratar en el análisis y para obtener una perspectiva más completa de los efectos de los diferentes tratamientos en el transcurso del tiempo (Stewart 1993, Pignon 1993, Jeng 1995, Clarke 1997).

11.6 Transformación de revisiones que utilizaron datos individuales de pacientes en formato de revisiones Cochrane

Debe estimularse la transformación en formato de revisiones Cochrane de revisiones relevantes ya existentes que han utilizado datos individuales de pacientes, a menos que pueda prepararse de otra manera una revisión Cochrane de mayor calidad. No obstante, estas transformaciones plantan desafíos particulares a los revisores y a los Grupos Colaboradores de Revisión (CRG).

Los metanálisis con DPI generalmente han sido realizados por grandes grupos colaboradores de investigadores de ensayos con DPI. A veces participan más de 100 personas incluyendo a los investigadores que proporcionan sus datos para el reanálisis, un secretariado de la organización, y en algunos casos, un comité asesor. Sin embargo, el tamaño de estos grupos, las políticas sociales implicadas y su publicación previa en formato en papel, pueden dificultar la adherencia con algunos procedimientos Cochrane referentes a recomendaciones de estilo y de formato. En particular:

- Para revisiones ya existentes con DPI, habitualmente no puede suministrarse retrospectivamente un protocolo y los CRG no requieren que haya uno antes de aceptar la revisión. Sin embargo, los revisores de DPI deben intentar enviar los protocolos de proyectos en curso durante sus primeras fases.

- El texto de la revisión con DPI habitualmente ha tenido muchas versiones en borrador y ha circulado entre todos los miembros del grupo colaborador para que sea comentado. No siempre es fácil llegar a un acuerdo en la redacción entre grupos tan grandes de manera que puede ser difícil cambiar el texto de una revisión para que sea incluido en la Cochrane Database of Systematic Reviews. Los editores y los revisores pares de los CRG deben ser comprensivos con respecto a esta limitación.

- Las etiquetas o Identificadores del Estudio habitualmente han sido seleccionados junto con los autores de los estudios y es poco probable que sea posible cambiarlos para que reflejen convenciones particulares.

- Para aquellas revisiones de DPI que ya han sido publicadas, el secretariado siempre debe haber obtenido suficiente información acerca de la contribución y el

consentimiento de la autoría con cada miembro del grupo colaborador para consentir la publicación de la revisión en una revista. Puede ser muy dispendioso obtener formatos Cochrane de contribución a la autoría con cada “autor”, de modo que el CRG debe aceptar de manera alternativa las declaraciones de autoría que los autores envían a la revista.

Los metanálisis con DPI deben ser revisados por pares durante el proceso normal de revisión por pares del CRG. Sin embargo los revisores pares deben tener claras las dificultades para hacer cambios (discutidas anteriormente). También deben tener en cuenta que las revisiones con DPI ya existentes, probablemente ya han sido sometidas a un exhaustivo proceso de revisión por pares antes de su presentación al CRG. Los manuscritos ya han sido examinados por el grupo de investigadores, por el secretariado y por el comité asesor de la revisión, e igualmente durante el proceso de revisión por pares de la revista en la cual la revisión con DPI ha sido publicada. Como sucede con todas las revisiones, la decisión final acerca de aceptar una revisión con DPI para su publicación como una revisión Cochrane en la Cochrane Database of Systematic Reviews debe ser tomada con el grupo editorial del CRG.

Aquellos CRGs que desean recibir asesoría con respecto a las revisiones con DPI, por ejemplo, con relación a su revisión por pares, deben contactar al Grupo de Métodos de metanálisis con DPI para recibir ayuda.

11.7 Metanálisis prospectivo

El metanálisis prospectivo es una forma especial de metanálisis con DPI. En este tipo de proyectos, un grupo de investigadores concuerda (antes de conocer los resultados de sus estudios) en combinar sus datos en el futuro. Se ha establecido un Grupo de Métodos en la Colaboración Cochrane para abordar este tema y para suministrar entrenamiento y apoyo en la realización de estos proyectos (apéndice 11b).

11. 6. Información adicional.

Muchos de los temas tratados aquí se amplían en: Stewart 1995 (Apéndice 11a). Ese informe también contiene ejemplos de cómo se han realizado anteriormente metanálisis con DPI, lo que quizá sea útil para los revisores que planifican uno actualmente. Los revisores de la Colaboración Cochrane a los que les interese información adicional pueden ponerse en contacto con el Grupo de Métodos.

Además, una presentación de diapositivas usada por el Grupo de Métodos en seminarios de formación está disponible en la página web de la Colaboración (http://www.cochrane.org/cochrane/training.htm).

11.9 References

Chalmers 1993. Chalmers I. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. Ann N Y Acad Sci 1993; 703: 156-65.

Clarke 1997. Clarke M, Stewart L. Individual patient data or published data meta-analysis: a systematic review [abstract]. *Proceedings of the Fifth Cochrane Collaboration Colloquium* 1997; 94, abstract 019.04.).

EBCTCG 1992. Early Breast Cancer Trialists’ Collaborative Group. Systematic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy: 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. Lancet 1992; 339:1-15, 71-85.

Jeng 1995. Jeng GT, Scott JR, Burmeister LF. A comparison of meta-analytic results using literature vs individual patient data: paternal cell immunization for recurrent miscarriage. JAMA 1995; 274: 830-6.

Pignon 1993. Pignon JP, Arriagada R. Meta-analysis. Lancet 1993; 341:964-5.

Stewart 1993. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data: is ther a difference? Lancet 1993; 341:418-22.

Stewart 1995. Stewart L, Clarke M, for the Cochrane Collaboration Working Group on meta-analyses using indivudal patient data. Practical methodology of meta-analyses (overviews) using updated individual patient data. Stat Med 1995; 14:2057-79.