Manual Cochrane de revisiones sistemáticas de intervenciones

Traducción a cargo del Centro Cochrane Iberoamericano de:

Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Versión 5.1.0

[actualizado en marzo de 2011]

Editores de la versión en inglés: Julian PT Higgins y Sally Green

ÍNDICE

INFORMACIÓN SOBRE EL *MANUAL* 3

Sobre el *Manual* 3

Cómo citar correctamente esta versión del *Manual* 4

Principales fuentes de apoyo 5

Agradecimientos 5

PARTE 1: REVISIONES COCHRANE 11

Capítulo 1: Introducción 12

Capítulo 2: Desarrollar una revisión Cochrane 21

Capítulo 3: Mantenimiento de las revisiones: actualizaciones, enmiendas y comentarios 42

Capítulo 4: Guía del contenido de un protocolo y una revisión Cochrane 66

PARTE 2: MÉTODOS GENERALES PARA LAS REVISIONES COCHRANE 103

Capítulo 5: Formular la pregunta de la revisión y desarrollar los criterios para incluir los estudios 104

Capítulo 6: La búsqueda de estudios 116

Capítulo 7: Selección de los estudios y obtención de los datos 165

Capítulo 8: Evaluación del riesgo de sesgo en los estudios incluidos 197

Capítulo 9: Análisis de los datos y realización de los metanálisis 256

Capítulo 10: Análisis del sesgo de informe 305

Capítulo 11: Presentación de los resultados y las tablas ‘resumen de los resultados’ 345

Capítulo 12: Interpretación de los resultados y obtención de las conclusiones 369

Parte 3: TEMAS ESPECIALES 397

Capítulo 13: Inclusión de estudios no aleatorizados 398

Capítulo 14: Efectos adversos 440

Capítulo 15: Incorporación de pruebas económicas 456

Capítulo 16: Temas especiales en estadística 488

Capítulo 17: Resultados notificados por el paciente 535

Capítulo 18: Revisiones de datos de pacientes individuales 551

Capítulo 19: Metanálisis prospectivo 566

Capítulo 20: Investigación cualitativa y revisiones Cochrane 579

Capítulo 21: Revisiones en salud pública y promoción de salud 599

Capítulo 22: Resúmenes de revisiones 614

INFORMACIÓN SOBRE EL *MANUAL*

Sobre el *Manual*

Manual Cochrane de revisiones sistemáticas de intervenciones

El *Manual Cochrane de revisiones sistemáticas de intervenciones* (el *Manual*) sirve de guía a los autores en la preparación de las revisiones Cochrane sobre intervenciones (incluidas las revisiones globales u *overviews* de revisiones).

Esta es la versión 5.1.0 del Manual, modificada por última vez en marzo de 2011. Incluye una serie de cambios para que coincida con la versión 5.1 del Review Manager. Los cambios se describen en el web del Manual.

La reimpresión de 2009 del libro (publicado por John Wiley & Sons, Ltd.) es la versión 5.0.2 del Manual; la impresión original (2008) del libro fue la versión 5.0.1 del Manual.

Actualizaciones

El *Manual* se actualiza periódicamente para incorporar los avances en la metodología de las revisiones sistemáticas y como respuesta a los comentarios de los usuarios. Para acceder a la versión más reciente, a los datos provisionales de la guía y a los detalles de versiones previas del *Manual*, por favor, consulte la página www.cochrane.org/training/cochrane-handbook

Se invita a los usuarios del *Manual* a que envíen comentarios y correcciones a los editores del *Manual*; los detalles de contacto se encuentran en la página web.

Reproducción y traducción

Se necesita obtener permiso para reproducir o traducir el *Manual Cochrane de revisiones sistemáticas de intervenciones*; por favor visite la página web.

Copyright © The Cochrane Collaboration

Cómo citar correctamente esta versión del *Manual*

Versión inglesa: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Versión en español: Centro Cochrane Iberoamericano, traductores. Manual Cochrane de Revisiones Sistemáticas de Intervenciones, versión 5.1.0 [actualizada en marzo de 2011] [Internet]. Barcelona: Centro Cochrane Iberoamericano; 2012. Disponible en http://www.cochrane.es/?q=es/node/269

Principales fuentes de apoyo

Actuales fuentes de apoyo

La Colaboración Cochrane
Medical Research Council, Reino Unido
Department of Health and Ageing, Australia
Monash University, Australia

Fuentes de apoyo anteriores

National Health Service Research and Development Programme, Reino Unido
Health Research Board, Irlanda
National Institute of Public Health, Noruega
Copenhagen Hospital Corporation, Dinamarca
Health Services Research and Development Service y la University of Texas Health Science Center, San Antonio, EE. UU.
US Veterans Health Administration, EE. UU.
Oxford Regional Health Authority, Reino Unido
Nuffield Provincial Hospitals Trust, Reino Unido
LW Frohlich Fund, EE. UU.
Norwegian Ministry of Health and Social Affairs, Noruega
Norwegian Research Council, Noruega
Glaxo Wellcome, Noruega

Agradecimientos

Estamos muy agradecidos a los miembros del Grupo Asesor Editorial de este Manual: Jane Cracknell, Toby Lasserson, Philippa Middleton, Jonathan Sterne y Lesley Stewart. Estamos igualmente agradecidos a todos los miembros del anterior Grupo Asesor por sus comentarios, y de forma preeminente, quisierámos dar las gracias a Doug Altman, Chris Cates, Mike Clarke, Jon Deeks, Donna Gillies, Andrew Herxheimer, Harriet MacLehose, Philippa Middleton, Ruth Mitchell, David Moher, Donald Patrick, Ian Shemilt, Lesley Stewart, Jessica Thomas, Jane Tierney y Danielle Wheeler.

Muchas personas contribuyeron con constructivas y oportunas revisiones por pares. Damos nuestro agradecimiento a Phil Alderson, Claire Allen, Judith Anzures, Chris Cates, Jonathan Craig, Miranda Cumpston, Chris Del Mar, Kay Dickersin, Christian Gluud, Peter Gøtzsche, Frans Helmerhorst, Jini Hetherington, Sophie Hill, Sally Hopewell, Steve McDonald, David Moher, Ann Møller, Duncan Mortimer, Karen New, Denise O'Connor, Jordi Pardo, Rob Scholten, Simon Thompson, Jan Vandenbroucke, Janet Wale, Phil Wiffen, Hywel Williams, Paula Williamson, Jim Wright y Diana Wyatt.

Jane Lean aportó un apoyo administrativo específico para la versión 5.0 del *Manual*. Además, el generoso y calificado apoyo administrativo y técnico de Claire Allen, Dave Booker, Jini Hetherington, Mónica Kjeldstrøm, Cindy Manukonga, Fiona Mathews, Rasmus Moustgaard, Jane Predl, Jacob Riis y Verena Roloff ha contribuido enormemente a la preparación y coordinación de este *Manual*. También queremos agradecer a Lucy Sayer, Fiona Woods y Laura Mellor y Jon Peacock de Wiley-Blackwell, su paciencia, apoyo y consejo, y también a Neil Manley por hacer un índice y a Wendy Langford por corregir las pruebas.

La versión 5 del *Manual* no hubiera sido posible sin el generoso apoyo a los editores por parte de sus colegas en el MRC Biostatistics Unit y el Institute of Public Health en Cambridge, Reino Unido y en el Australasian Cochrane Centre, Monash University, Australia.

Agradecimientos de la versión en español

Queremos agradecer a José Luis Casasbuenas, María Teresa Gil del Real y muy especialmente a Juan C. Vázquez su apoyo en la traducción del Manual, así como a Aleix Bonfill, Xavier Bonfill, Marta Gorgues, Marta Roqué, Ivan Solà, Gerard Urrútia y Sera Tort sus aportaciones durante la revisión de la traducción.

Principales contribuidores

Ver también el Capítulo 1 (Sección 1.4) para obtener una lista de los principales contribuidores de las versiones anteriores.

Editores del *Manual*

Julian Higgins es *senior statistician* en la MRC Biostatistics Unit, Institute of Public Health, University of Cambridge. También es *visiting fellow* en el Cochrane Centre de Oxford, Reino Unido.

Sally Green es *professorial fellow* en el Institute of Health Services Research en Monash University, Melbourne, Australia, y directora del Australasian Cochrane Centre.

Autores de capítulo y editores

Acquadro, Catherine
Mapi Research Institute
Lyon, Francia

Alderson, Philip
National Institute for Health and Clinical Excellence
London/Manchester
Reino Unido

Altman, Douglas G
Centre for Statistics in Medicine
University of Oxford
Oxford, Reino Unido

Armstrong, Rebecca
The McCaughey Centre: VicHealth Centre for the Promotion of Mental Health and Community Wellbeing
University of Melbourne
Melbourne
Australia

Askie, Lisa M
NHMRC Clinical Trials Centre
University of Sydney
Camperdown
Australia

Becker, Lorne A
Department of Family Medicine
SUNY Upstate Medical University
Syracuse, NY
Estados Unidos

Mallender, Jacqueline
Matrix Knowledge Group Ltd.
London, Reino Unido

Manheimer, Eric
Center for Integrative Medicine
University of Maryland School of Medicine
Baltimore, MA, Estados Unidos

McDaid, David
Personal Social Services Research Unit
London School of Economics and Political Science
London, Reino Unido

Moher, David
Chalmers Research Group, Children's Hospital of Eastern Ontario Research Institute;
Department of Epidemiology and Community Medicine, University of Ottawa
Ottawa
Canadá

Mugford, Miranda
Health Economics Group
School of Medicine, Health Policy and Practice
University of East Anglia
Norwich, Reino Unido

Mulrow, Cynthia D
University of Texas Health Science Center at San Antonio
San Antonio, TX
Estados Unidos

Berlin, Jesse A
Pharmacoepidemiology
Johnson & Johnson Pharmaceutical Research and Development
Titusville, NJ
Estados Unidos

Booth, Andrew
School of Health and Related Research
University of Sheffield
Sheffield
Reino Unido

Byford, Sarah
Centre for the Economics of Mental Health
Institute of Psychiatry
King's College
London
Reino Unido

Clarke, Mike
UK Cochrane Centre
National Institute for Health Research
Oxford
Reino Unido

Deeks, Jonathan J
Department of Public Health and Epidemiology
University of Birmingham
Birmingham
Reino Unido

Doyle, Jodie
The McCaughey Centre: VicHealth Centre for the Promotion of Mental Health and Community Wellbeing
University of Melbourne
Melbourne
Australia

Drummond, Michael
Centre for Health Economics
University of York
York, Reino Unido

Egger, Matthias
Institute of Social and Preventive Medicine
University of Bern
Bern
Suiza

Eisenstein, Eric
Duke Clinical Research Center
Duke University
Durham, NC
Estados Unidos

Noyes, Jane
Centre for Health-Related Research
School of Healthcare Sciences
Bangor University
Bangor
Wales
Reino Unido

O'Connor, Denise
Australasian Cochrane Centre
Monash University
Melbourne
Australia

Oxman, Andrew D
Preventive and International Health Care Unit
Norwegian Knowledge Centre for the Health Services
Oslo
Noruega

Patrick, Donald L
Department of Health Services and Seattle Quality of Life Group
University of Washington
Seattle, WA
Estados Unidos

Pearson, Alan
Joanna Briggs Institute
University of Adelaide
Adelaide
Australia

Popay, Jennie
Institute for Health Research
Lancaster University
Lancaster
Reino Unido

Price, Deirdre
Department of Clinical Pharmacology
University of Oxford
Oxford
Reino Unido

Reeves, Barnaby C
Bristol Heart Institute
University of Bristol
Bristol
Reino Unido

Scholten, Rob JPM
Dutch Cochrane Centre
Academic Medical Center
Amsterdam
Holanda

Ghersi, Davina
Department of Research Policy and Cooperation
World Health Organization
Geneva
Suiza

Glanville, Julie
Centre for Reviews and Dissemination
University of York
York
Reino Unido

Glasziou, Paul P
Department of Primary Health Care
University of Oxford
Oxford
Reino Unido

Green, Sally
Australasian Cochrane Centre
Monash University
Melbourne
Australia

Guyatt, Gordon H
Departments of Clinical Epidemiology and Biostatics
McMaster University
Ontario
Canadá

Hannes, Karin
Belgian Centre for Evidence-Based Medicine
Leuven
Bélgica

Herxheimer, Andrew
Co-founder, DIPEx; Emeritus Fellow, UK Cochrane Centre
London
Reino Unido

Higgins, Julian PT
MRC Biostatistics Unit
Cambridge
Reino Unido

Knapp, Martin
Institute of Psychiatry
King's College London; and
London School of Economics
London
Reino Unido

Schünemann, Holger J
INFORMA/CLARITY Research/Department of Epidemiology
National Cancer Institute Regina Elena
Rome
Italia

Shemilt, Ian
Health Economics Group
School of Medicine, Health Policy and Practice
University of East Anglia
Norwich
Reino Unido

Sterne, Jonathan AC
Department of Social Medicine
University of Bristol
Bristol
Reino Unido

Stewart, Lesley A
Centre for Reviews and Dissemination
University of York
York
Reino Unido

Tierney, Jayne F
MRC Clinical Trials Unit
London
Reino Unido

Vale, Luke
Health Economics Research Unit
University of Aberdeen
Aberdeen
Reino Unido

Vist, Gunn E
Preventive and International Health Care Unit
Norwegian Knowledge Centre for the Health Services
Oslo
Noruega

Walker, Damian
Health Systems Program
Department of International Health
Johns Hopkins Bloomberg School of Public Health
Baltimore, MA
Estados Unidos

Lefebvre, Carol
UK Cochrane Centre
National Institute for Health Research
Oxford
Reino Unido

Loke, Yoon K
School of Medicine,
Health Policy and Practice
University of East Anglia
Norwich
Reino Unido

Waters, Elizabeth
The McCaughey Centre: VicHealth Centre for the Promotion of Mental Health and Community Wellbeing
University of Melbourne
Melbourne
Australia

Wells, George A
Department of Epidemiology and Community Medicine
University of Ottawa
Ottawa
Ontario
Canadá

PARTE 1: REVISIONES COCHRANE

ÍNDICE
CAPÍTULO 1: INTRODUCCIÓN

PUNTOS CLAVE

1.1 LA COLABORACIÓN COCHRANE

1.1.1 Introducción

Recuadro 1.1.a: Los principios de la Colaboración Cochrane

1.1.2 La estructura de la Colaboración Cochrane

1.1.3 Publicación de las revisiones Cochrane

Recuadro 1.1.b: Bases de datos publicadas en la Biblioteca Cochrane

1.2 REVISIONES SISTEMÁTICAS

1.2.1 La necesidad de revisiones sistemáticas

1.2.2 ¿Qué es una revisión sistemática?

1.3 ACERCA DE ESTE *MANUAL*

1.4 COLABORADORES DEL *MANUAL*

1.5 INFORMACIÓN SOBRE EL CAPÍTULO

1.6 REFERENCIAS

CAPÍTULO 1: INTRODUCCIÓN

Autores: Sally Green, Julian PT Higgins, Philip Alderson, Mike Clarke, Cynthia D Mulrow y Andrew D Oxman.

Puntos clave

- Las revisiones sistemáticas tienen como objetivo reunir toda evidencia que se corresponda con unos criterios de elegibilidad establecidos previamente, con el fin de orientar un tema específico de investigación.
- El propósito de las revisiones sistemáticas es minimizar sesgos mediante la aplicación de métodos sistemáticos y explícitos.
- La Colaboración Cochrane prepara, mantiene y fomenta la difusión de revisiones sistemáticas para fundamentar decisiones relacionadas con la atención sanitaria (revisiones Cochrane).
- Las revisiones Cochrane se publican en la *Base de Datos Cochrane de Revisiones Sistemáticas*, publicada en *The Cochrane Library*.
- *El Manual Cochrane de revisiones sistemáticas de intervenciones* contiene una guía metodológica para la preparación y actualización de las revisiones Cochrane sobre intervenciones y de las revisiones Cochrane globales.

1.1 La Colaboración Cochrane

1.1.1 Introducción

La Colaboración Cochrane (www.cochrane.org) es una organización internacional cuyo propósito principal es ayudar a tomar decisiones sanitarias bien fundamentadas preparando, manteniendo y promocionando el acceso a las revisiones sistemáticas basadas en la evidencia. Al aportar una síntesis fiable de la evidencia disponible sobre un tema determinado, las revisiones sistemáticas se adhieren al precepto de que la ciencia es acumulativa, y facilita las decisiones al tener en cuenta toda la evidencia del efecto de una intervención. Desde que fue fundada en 1993, la Colaboración Cochrane ha crecido hasta alcanzar más de 15.000 participantes en más de 100 países, llegando a ser así la organización más grande dedicada a una labor como ésta (Allen 2006, Allen 2007). La Colaboración internacional se inició un año después del establecimiento del Centro Cochrane en Oxford, Reino Unido, fundado por Sir Ian Chalmers y algunos colegas, y bautizado después con el nombre del epidemiólogo británico Archie Cochrane. La Colaboración Cochrane es hoy una iniciativa internacionalmente reconocida (Clarke 2005, Green 2005).

El trabajo de la Colaboración Cochrane está basado en un conjunto de 10 principios básicos, enumerados en el Recuadro 1.1.a.

Recuadro 1.1.a: Los principios de la Colaboración Cochrane

1. *Colaboración*, promoviendo una buena comunicación, tanto externa como interna, la toma de decisiones abierta y el trabajo en equipo.

2. *Basar su desarrollo en el entusiasmo individual* involucrando y dando apoyo a personas con formaciones diferentes, con experiencias diversas y procedencias diferentes.

3. *Evitar la duplicación del trabajo* mediante una buena gestión y coordinación para llegar a la máxima economía de esfuerzos.

4. *Minimizar los sesgos* mediante una variedad de estrategias, como son el rigor científico, la participación más amplia posible o evitando conflictos de interés.

5. *Actualización permanente* de las revisiones Cochrane, asumiendo el compromiso de asegurar que éstas identificarán e incorporarán la nueva evidencia que vaya apareciendo.

6. *Esforzarse por producir material relevante*, fomentando la evaluación de las intervenciones sanitarias a partir de resultados que sean clave para quienes deben tomar decisiones en el campo de la salud.

7. *Promover el acceso* mediante la amplia divulgación de la producción de la Colaboración, aprovechando alianzas estratégicas y asegurando una política de precios, un contenido y unos medios adecuados para cubrir las necesidades de los usuarios en todo el mundo.

8. *Asegurar la calidad*, al estar abierta y receptiva a la crítica, aplicar los adelantos de índole metodológica y desarrollar sistemas para mejorar la calidad.

9. *Continuidad*, al asegurar que se mantenga y renueve la responsabilidad de los revisores, los procesos editoriales y las funciones clave.

10. *Facilitar una amplia participación* en las actividades de la Colaboración al minimizar obstáculos para las contribuciones y al estimular la diversidad.

1.1.2 La estructura de la Colaboración Cochrane

Las actividades de la Colaboración Cochrane se organizan alrededor de cincuenta y un Grupos de Revisión Cochrane (*Cochrane Review Groups*, CRG), responsables de preparar y actualizar las revisiones dentro de áreas específicas de la atención sanitaria. Los miembros de estos grupos incluyen investigadores, profesionales sanitarios y personas que utilizan los servicios de salud (usuarios y/o pacientes), que comparten todos un entusiasmo común, para producir evidencia fiable, actualizada y relacionada con la prevención y el tratamiento de problemas o grupos de problemas específicos de salud.

Los Grupos de Revisión Cochrane reciben apoyo para el desarrollo de revisiones por parte de los Grupos de Metodología, Centros y Ámbitos. Los Grupos de Metodología Cochrane (*Cochrane Methods Groups*) proporcionan un foro para que estos expertos discutan el desarrollo, la evaluación y la aplicación de los métodos utilizados para preparar las revisiones Cochrane. Juegan un importante papel en la producción del *Manual Cochrane de Revisiones Sistemáticas de Intervenciones* (el *Manual*) cuyos capítulos contienen información acerca de los Grupos de Métodos relacionados. Los Centros Cochrane están localizados en diferentes países y juntos representan todas las regiones, y proveen formación y apoyo para los autores de la revisión y CRG además de fomentar y promocionar el acceso a las revisiones Cochrane. Los Ámbitos Cochrane (*Cochrane Fields*) se ocupan de otras dimensiones sanitarias más allá de los problemas de salud, tales como el escenario de la atención (por ejemplo, atención primaria), el tipo de consumidor (por ejemplo, niños), o el tipo de intervención (por ejemplo, vacunas). Las personas vinculadas a los Ámbitos se preocupan de asegurar que las prioridades y perspectivas en su ámbito de interés quedan reflejadas en el trabajo de los CRG.

1.1.3 Publicación de las revisiones Cochrane

Las revisiones Cochrane se publican en línea a texto completo en la *Base de Datos Cochrane de Revisiones Sistemáticas* (*Cochrane Database of Systematic Reviews*, CDSR, sigla en inglés), que es un componente básico de *The Cochrane Library*. *The Cochrane Library* está publicada por Wiley-Blackwell en Internet (www.thecochranelibrary.com) y en CD-ROM, y es accesible de forma gratuita en algunos países gracias a licencias nacionales y en la mayoría de los lugares de menos recursos gracias al acceso libre que provee Wiley-Blackwell. En otros lugares se requiere suscripción o *pay-per-view*. Además de la CDSR, *The Cochrane Library* contiene otra serie de fuentes de información, enumeradas en el Recuadro 1.1b.

La CDSR se publica doze veces al año, cada vez con nuevas revisiones y actualizaciones de las revisiones existentes. El volumen 3, 2011 de la CDSR (Marzo 2011) contiene más de 4.500 revisiones Cochrane y cerca de 2.000 protocolos de revisiones en curso.

Recuadro 1.1.b: Bases de datos publicadas en la Biblioteca Cochrane

- *La Base de Datos Cochrane de Revisiones Sistemáticas (Cochrane Database of Systematic Reviews, CDSR) contiene el texto completo (que incluye métodos, resultados y conclusiones) de las revisiones y protocolos Cochrane.*

- *La Base de Datos de Resúmenes de Revisiones de Efectos (Database of Abstracts of Reviews of Effects, DARE) compilada y mantenida por el Centro de Revisiones y Divulgación (Centre for Reviews and Dissemination) de York, Reino Unido, incluye evaluaciones críticas y resúmenes estructurados de otras revisiones sistemáticas que cumplen unos criterios de calidad explícitos.*

- *El Registro Central Cochrane de Ensayos Controlados (The Cochrane Central Register of Controlled Trials, CENTRAL) contiene información bibliográfica de cientos de miles de ensayos controlados, e incluye también informes publicados en resúmenes de conferencias y otras fuentes no incluidas en otras bases de datos bibliográficas.*

- *El Registro Cochrane de Metodología (Cochrane Methodology Register) contiene información bibliográfica de artículos y libros en el campo de la investigación en revisiones, así como un registro prospectivo de estudios metodológicos*

- *La sección de la Colaboración Cochrane* de *The Cochrane Library* incluye los detalles de contacto y otros datos sobre los CRG y otras entidades que participan en la Colaboración Cochrane.

1.2 Revisiones sistemáticas

1.2.1 La necesidad de revisiones sistemáticas

Los proveedores, los usuarios y pacientes, los investigadores y todas aquellas personas que elaboran políticas sanitarias se enfrentan a cantidades inmanejables de información relativa a evidencia en investigación sanitaria. Es poco probable que todos dispongan del tiempo, las habilidades y los recursos para identificar, evaluar e interpretar esta evidencia e incorporarla a las decisiones sanitarias. Las revisiones Cochrane responden a este reto identificando, evaluando y sintetizando la evidencia basada en la investigación y presentándola en un formato accesible (Mulrow 1994).

1.2.2 ¿Qué es una revisión sistemática?

Una revisión sistemática tiene como objetivo reunir toda la evidencia empírica que cumple unos criterios de elegibilidad previamente establecidos, con el fin de responder una pregunta específica de investigación. Utiliza métodos sistemáticos y explícitos, que se eligen con el fin de minimizar sesgos, aportando así resultados más fiables a partir de los cuales se puedan extraer conclusiones y tomar decisiones (Antman 1992, Oxman 1993). Los elementos fundamentales de una revisión sistemática son:

- *Un conjunto de objetivos claramente establecidos, con criterios de elegibilidad de estudios previamente definidos.*
- *Una metodología explícita y reproducible.*
- *Una búsqueda sistemática que identifique todos los estudios que puedan cumplir los criterios de elegibilidad.*
- *Una evaluación de la validez de los resultados de los estudios incluidos, por ejemplo mediante la evaluación del riesgo de sesgos.*
- *Una presentación sistemática y una síntesis de las características y resultados de los estudios incluidos.*

Muchas de las revisiones sistemáticas contienen metanálisis. El metanálisis consiste en la aplicación de métodos estadísticos para resumir los resultados de estudios independientes (Glass 1976). Al combinar la información de todos los estudios relevantes, el metanálisis puede obtener estimaciones más precisas de los efectos sobre la atención sanitaria que las derivadas de los estudios individuales incluidos en una revisión (ver el Capítulo 9, Sección 9.1.3). También permite investigar la consistencia de la evidencia entre estudios y explorar las diferencias entre estudios.

1.3 Acerca de este *Manual*

La ciencia de la síntesis de la investigación evoluciona rápidamente; de aquí que los métodos empleados en el desarrollo de revisiones Cochrane hayan evolucionado con el tiempo. La intención del *Manual Cochrane de revisiones sistemáticas de intervenciones* (el *Manual*) es ayudar a los autores de revisiones Cochrane a tomar decisiones apropiadas acerca de los métodos que emplean, más que dictar normas arbitrariamente. En la medida de lo posible, las recomendaciones se complementan con evidencia empírica. Las guías que aquí se aportan intentan ayudar a los autores de revisiones a ser sistemáticos, a estar informados y a ser explícitos (pero no mecanicistas) en el planteamiento de preguntas y en el proceso de encontrar respuestas a esas preguntas.

La interpretación e implementación de esta guía requiere criterio y debe hacerse en conjunto con las bases editoriales de los CRG.

Este *Manual* se centra en las revisiones sistemáticas sobre los efectos de las intervenciones. La mayoría de las recomendaciones que contiene están orientadas a la síntesis de los ensayos clínicos y de los ensayos aleatorizados en particular, ya que aportan evidencia más fiable que otros diseños de estudios sobre los efectos de las intervenciones sanitarias (Kunz 2007). Sin embargo, algunos capítulos añaden la recomendación de incluir otros tipos de evidencia, particularmente en prácticas sanitarias en las que los ensayos aleatorizados puede que no sean posibles o apropiados, y en consideraciones de seguridad o efectos adversos. En el 2003, la Colaboración Cochrane amplió su campo de acción para incluir las revisiones Cochrane de validez de pruebas diagnósticas. Las guías para el manejo de estas revisiones se encuentran en un documento separado: el *Manual Cochrane de revisiones sistemáticas de validez de pruebas diagnósticas.*

Este *Manual* contiene 22 capítulos organizados en tres partes. La parte 1 presenta las revisiones Cochrane, incluye su planificación y preparación, su mantenimiento y actualización y termina con una guía de los contenidos de una revisión o protocolo Cochrane. La parte 2 proporciona una guía metodológica general relativa a todas las revisiones Cochrane, tratando la formulación de la pregunta, criterios de elegibilidad, búsqueda, extracción de datos, sesgos intra-estudio, análisis de los datos, sesgos del informe, presentación e interpretación de resultados. La parte 3 trata temas especiales relativos a algunas, pero no a todas, las revisiones Cochrane, que incluyen aspectos particulares sobre efectos adversos, metanálisis de diseños de estudios no habituales y metanálisis con datos de pacientes individuales. Esta parte incluye capítulos que tratan de la inclusión en las revisiones de evaluaciones económicas, estudios no aleatorizados, investigación cualitativa, resultados informados por el paciente, metanálisis prospectivos y revisiones en la promoción sanitaria y salud pública. Un capítulo final describe un nuevo tipo de revisión, las revisiones globales (*Overviews of reviews*).

Cada capítulo contiene una lista de puntos clave que resumen la información y presentan los mensajes clave destinados a los autores de revisiones.

El *Manual* ha sido desarrollado en gran parte por los Grupos de Metodología de la Colaboración Cochrane, cuyos miembros llevan a cabo gran parte de la investigación metodológica y empírica en que se basa la guía.

Aunque los principales usuarios previstos para el *Manual* son los autores de las revisiones Cochrane sobre intervenciones, muchos de los principios y métodos se pueden aplicar a revisiones sistemáticas utilizadas en otros tipos de investigación y a revisiones sistemáticas realizadas por otros investigadores (Moher 2007).

1.4 Colaboradores del *Manual*

"Si he podido ver lejos es porque estaba en hombros de gigantes"

Isaac Newton

Este *Manual Cochrane de revisiones sistemáticas de intervenciones* (versión 5) es una valiosa revisión de un documento que ha evolucionado con el tiempo desde los primeros días de la Colaboración Cochrane. Muchos capítulos se basan en versiones previas del *Manual*, mientras otros han sido escritos por primera vez para esta versión 5. Es un verdadero esfuerzo de colaboración, que refleja los principios de la Colaboración Cochrane. Muchas personas han contribuido directamente a esta revisión, ejerciendo como autores de capítulos, editores de capítulos, revisores por pares, miembros del Grupo Asesor del Manual Cochrane, y de muchas otras maneras. El *Manual* también refleja las inapreciables contribuciones de editores anteriores, miembros anteriores y actuales de los Grupos Cochrane de Metodología, autores de revisiones, Grupos Cochrane de Revisión, el Grupo Asesor de RevMan, Centros Cochrane y Ámbitos Cochrane.

La guía metodológica inicial para autores de revisión Cochrane fue desarrollada 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. Fue publicada en marzo de 1994 como *Sección VI: Preparación y mantenimiento de las revisiones sistemáticas ('La Caja de Herramientas de la Colaboración Cochrane')*, de un amplio manual para la Colaboración. Describía el formato original estructurado de una revisión Cochrane, que fue desarrollado por Mike Clarke, Murray Enkin, Chris Silagy y Mark Starr, con la aportación de muchos otros. La guía llegó a ser un documento autónomo en octubre de 1996 como el *Manual de la Colaboración Cochrane* (versión 3), bajo la dirección editorial de Andy Oxman y Cynthia Mulrow, apoyados por el recientemente formado Grupo Asesor del Manual. La versión 4, denominada el *Manual Cochrane de Revisores*, salió en 1999 coincidiendo con el lanzamiento de RevMan 4 y fue editado por Mike Clarke y Andy Oxman desde 1999 hasta diciembre de 2003, cuando Phil Alderson, Julian Higgins y Sally Green se hicieron editores (desde la versión 4.2.1). La inclusión de las revisiones Cochrane de validez de las pruebas diagnósticas y la necesidad de un nuevo manual específico para ellos impulsó, desde la versión 4.2.4 de marzo de 2005, el cambio del nombre a *Manual Cochrane de revisiones sistemáticas de intervenciones*, editado por Julian Higgins y Sally Green.

Los editores actuales del *Manual* reciben apoyo del Consejo Asesor del Manual. Este Consejo Asesor del Manual está compuesto actualmente por: Lisa Askie, Chris Cates, Jon Deeks, Matthias Egger, Davina Ghersi, Donna Gillies, Paul Glasziou, Sally Green (coordinadora), Andrew Herxheimer, Julian Higgins (coordinador), Jane Lane (administración), Carol Lefebvre, Harriet MacLehose, Philippa Middleton, Ruth Mitchell, David Moher, Miranda Mugford, Jane Noyes, Donald Patrick, Jennie Popay, Barney Reeves, Jacob Riis, Ian Shemilt, Jonathan Sterne, Lesley Stewart, Jessica Thomas, Jayne Tierney y Danielle Wheeler.

Además de los actuales editores enumerados anteriormente, los mencionados a continuación hicieron importantes contribuciones a versiones previas del *Manual*: Christina Aguilar, Doug Altman, Bob Badgett, Hilda Bastian, Lisa Bero, Michael Brand, Joe Cavellero, Mildred Cho, Kay Dickersin, Lelia Duley, Frances Fairman, Jeremy Grimshaw, Gord Guyatt, Peter Gøtzsche, Jeph Herrin, Nicki Jackson, Monica Kjeldstrøm, Jos Kleijnen, Kristen Larson, Valerie Lawrence, Eric Mamheimer, Rasmus Moustgaard, Melissa Ober, Drummond Rennie, Dave Sackett, Mark Starr, Nicola Thornton, Luke Vale y Veronica Yank.

1.5 Información sobre el capítulo

Autores: Sally Green, Julian PT Higgins, Philip Alderson, Mike Clarke, Cynthia D Mulrow y Andrew D Oxman.

La versión en inglés de este capítulo debe ser citada como: Green S, Higgins JPT, Alderson P, Clarke M, Mulrow CD, Oxman AD. Chapter 1: Introduction. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

1.6 Referencias

Allen 2006
Allen C, Clarke M. International activity in Cochrane Review Groups with particular reference to China. Chinese Journal of Evidence-based Medicine. 2006;6:541-545.

Allen 2007
Allen C, Clarke M, Tharvan P. International activity in Cochrane Review Groups with particular reference to India. The National Medical Journal of India. 2007;20:250-255.

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;268:240-248.

Clarke 2005
Clarke M. Cochrane Collaboration. In: Armitage P, Colton T (editors). Encyclopedia of Biostatistics (2nd edition). Chichester (UK): John Wiley & Sons; 2005.

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

Green 2005
Green S, McDonald S. The Cochrane Collaboration: More than systematic reviews? Internal Medicine Journal. 2005;35:4-5.

Kunz 2007
Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. Cochrane Database of Systematic Reviews 2007, Issue 2. Art No: MR000012.

Moher 2007
Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLOS Medicine. 2007;4:e78.

Mulrow 1994
Mulrow CD. Rationale for systematic reviews. BMJ. 1994;309:597-599.

Oxman 1993
Oxman AD, Guyatt GH. The science of reviewing research. Annals of the New York Academy of Sciences. 1993;703:125-133.

ÍNDICE
CAPÍTULO 2: DESARROLLAR UNA REVISIÓN COCHRANE

PUNTOS CLAVE

2.1 JUSTIFICACIÓN DE LOS PROTOCOLOS

2.2 FORMATO DE UNA REVISIÓN COCHRANE

2.2.1 Justificación del formato de una revisión Cochrane
2.2.2 Esquema de un protocolo de revisión Cochrane
Recuadro 2.2.a: Secciones de un protocolo en una revisión Cochrane
2.2.3 Esquema detallado de una revisión Cochrane
Recuadro 2.2.b: Secciones de una revisión sistemática Cochrane

2.3 ASPECTOS LOGÍSTICOS PARA HACER UNA REVISIÓN

2.3.1 Motivación para asumir una revisión
2.3.2 Planificar el tema y el alcance de una revisión
2.3.3 Registrar un protocolo
2.3.4 El equipo de revisión
2.3.4.1 La importancia de un equipo
2.3.4.2 Participación de los pacientes y usuarios
2.3.4.3 Grupos asesores
Recuadro 2.3.a: Un ejemplo de los beneficios de utilizar un grupo asesor en el proceso de planificación
2.3.5 Software Cochrane para autores de revisión y apoyo editorial a los Grupos Cochrane de Revisión
2.3.6 Formación
2.3.7 Procedimientos editoriales de un Grupo Cochrane de Revisión
2.3.8 Recursos para una revisión sistemática
Recuadro 2.3.b: Calendario para una revisión Cochrane

2.4 PUBLICACIÓN DE REVISIONES COCHRANE EN REVISTAS BIOMÉDICAS Y LIBROS

2.5 PUBLICACIÓN DE REVISIONES PUBLICADAS PREVIAMENTE COMO REVISIONES COCHRANE

2.6 DECLARACIÓN DE INTERÉS Y PATROCINIO COMERCIAL

Recuadro 2.6.a: Código de Conducta de la Colaboración Cochrane para evitar posibles conflictos de interés financiero

2.7 INFORMACIÓN DEL CAPÍTULO

2.8 REFERENCIAS

CAPÍTULO 2: DESARROLLAR UNA REVISIÓN COCHRANE

Editores: Sally Green y Julian PT Higgins.

Puntos clave

- La publicación de protocolos de revisiones Cochrane en la *Base de Datos Cochrane de Revisiones Sistemáticas (CDSR* en su sigla en inglés*),* previa a la publicación de la revisión Cochrane, reduce el impacto de sesgos de los autores, fomenta la transparencia de los métodos y los procesos planteados, disminuye la posibilidad de duplicación, y permite la revisión por pares de los métodos planificados;

- Las revisiones Cochrane y los protocolos de revisiones se preparan en el software de la Colaboración Cochrane *Review Manager* (RevMan), y tienen un formato único;

- En este capítulo se incluyen las secciones básicas de una revisión Cochrane de Intervención;

- Los títulos de las revisiones Cochrane de Intervención son acordados y registrados con los Grupos Cochrane de Revisión (*CRG* en su sigla en inglés), encargados del proceso editorial de la publicación de los protocolos y las revisiones;

- Las revisiones Cochrane las preparadas equipos de revisores;

- Existen guías para la co-publicación de las revisiones Cochrane en otras revistas científicas;

- La Colaboración Cochrane tiene un código de conducta para evitar posibles conflictos de interés financiero.

2.1 Justificación de los protocolos

Preparar una revisión Cochrane es complejo e implica muchos juicios. Con el fin de disminuir el riesgo potencial de sesgos en el proceso de revisión, estos juicios deben hacerse de forma que no dependan de los hallazgos de los estudios incluidos en la revisión. El conocimiento previo de los resultados de un estudio potencialmente elegible por parte de los autores de la revisión puede, por ejemplo, influenciar la definición de una pregunta de revisión sistemática, el criterio subsecuente de la elegibilidad del estudio, la elección de las comparaciones a evaluar entre intervenciones, o las variables de resultado a publicar en la revisión. Como las revisiones Cochrane son por naturaleza retrospectivas (una excepción es el metanálisis prospectivo, como se describe en el capítulo 19), es importante que los métodos que se van a utilizar sean establecidos y documentados previamente. La publicación de un protocolo de una revisión sistemática, previa al conocimiento de los estudios disponibles, reduce el impacto de los sesgos de los autores de la revisión, estimula la transparencia de los métodos y los procesos planteados, disminuye la posibilidad de duplicación, y permite la revisión por pares de los métodos propuestos (Light 1984).

Aunque la intención debe ser mantenerse fiel al protocolo de revisión publicado, a veces los cambios en estos protocolos son necesarios. Esta situación es similar a la de un protocolo para un ensayo aleatorizado, que algunas veces debe ser cambiado para adaptarlo a circunstancias imprevistas, como problemas de inclusión de los participantes, la extracción de datos, o la incidencia de eventos inesperados. Aunque debe hacerse el esfuerzo de permanecer fiel al protocolo predeterminado, esto no siempre es posible o apropiado. Sin embargo, es importante que los cambios en el protocolo no se hagan basándose en cómo éstos afectan el resultado del estudio de investigación. Las decisiones tomadas a posteriori, cuando el impacto en los resultados de la investigación ya se conoce, tales como excluir estudios seleccionados de una revisión sistemática, son altamente susceptibles de sesgos y deben evitarse.

Los protocolos de revisiones Cochrane se publican antes de terminar la revisión sistemática en la *Base de Datos Cochrane de Revisiones Sistemáticas (CDSR).* Los cambios en el protocolo deben documentarse e informarse en la sección 'Diferencias entre el protocolo y la revisión' de la revisión terminada, y los análisis de sensibilidad (ver Capítulo 9, Sección 9.7) que exploran el impacto de las desviaciones del protocolo deben asumirse cuando sea posible.

2.2 Formato de una revisión Cochrane

2.2.1 Justificación del formato de una revisión Cochrane

Todas las revisiones Cochrane de Intervención tienen el mismo formato. Los beneficios de este formato uniforme incluyen:

1. Ayudar a los lectores a encontrar rápidamente los resultados de la investigación y a evaluar la validez, la aplicabilidad y las implicaciones de esos resultados;

2. Guiar a los autores de la revisión a informar sobre su trabajo de forma explícita y concisa, y minimizar el esfuerzo requerido para hacerlo;

3. Facilitar la publicación electrónica y el mantenimiento de las revisiones;

4. Posibilitar el desarrollo de productos derivados (Revisiones globales u *overview* de revisiones, ver Capítulo 22) y estudios empíricos basados en múltiples revisiones sistemáticas.

El formato es suficientemente flexible para acomodarse a diferentes tipos de revisiones, incluidas las que hacen una única comparación, las que hacen múltiples comparaciones y las que analizan datos de pacientes individuales.

Los encabezamientos y las tablas incluidas en el RevMan guían a los autores cuando están preparando su revisión y facilitan a los lectores la identificación de la información que es de su particular interés. La lista de los encabezamientos disponibles en el RevMan aparece en la Sección 2.2.2 y 2.2.3. Una guía detallada del contenido que debe seguir a cada encabezamiento se encuentra en el Capítulo 4.

2.2.2 Esquema de un protocolo de revisión Cochrane

El recuadro 2.2.a enumera los elementos que definen un protocolo completo de una revisión Cochrane, e indica cómo debe aparecer el protocolo en la *CDSR* (que no tiene por qué ser igual que en el RevMan). Si alguna de las secciones marcadas con un asterisco (*) aparece vacía, el protocolo no será publicado hasta que no se haya agregado contenido en la sección, es decir, son 'campos obligatorios'.

Recuadro 2.2.a: Secciones de un protocolo en una revisión Cochrane

- Título*
- Información del protocolo:
  - Autores*
  - Persona de contacto*
  - Fechas
  - Novedades
  - Historial
- Protocolo:
  - Antecedentes*
  - Objetivos*
  - Métodos:
    - Criterios de selección de los estudios:
      - Tipos de estudios*
      - Tipos de participantes*
      - Tipos de intervenciones*
      - Tipos de medidas de desenlace*

- Métodos de búsqueda para la identificación de los estudios*
- Extracción y análisis de los datos*
- Agradecimientos
- Referencias:
  - Otras referencias:
    - Referencias adicionales
    - Otras versiones publicadas de esta revisión
- Tablas y figuras:
  - Tablas adicionales
  - Figuras

Información complementaria:

- Apéndices
- Comentarios:
  - Título
  - Resumen
  - Respuesta
  - Contribuciones

Sobre el artículo:

- Contribuciones de los autores
- Declaraciones de interés*
- Financiación:
  - Fuentes internas
  - Fuentes externas
- Notas publicadas

2.2.3 Esquema detallado de una revisión Cochrane

El recuadro 2.2.b enumera los elementos que definen una revisión sistemática Cochrane completa, e indica cómo debe aparecer la revisión en la *CDSR* (que no tiene por qué ser la misma que en el RevMan). Si alguna de las secciones marcadas con un asterisco (*) aparece vacía, la revisión no será publicada hasta que se agregue contenido en esa sección, es decir, son 'campos obligatorios'.

Recuadro 2.2.b: Secciones de una revisión sistemática Cochrane

Título*

Información de la revisión:

- Autores*
- Persona de contacto*
- Fechas*
- Novedades
- Historial

Resumen:

- Antecedentes*
- Objetivos*
- Estrategia de búsqueda*
- Extracción y análisis de los datos*
- Resultados*
- Conclusiones de los autores*

Resumen en términos sencillos:

- Título en términos sencillos*
- Texto resumido*

La revisión:
- Antecedentes*
- Objetivos*
- Métodos:
  - Criterios de selección de estudios para esta revisión:
    - Tipos de estudios*
    - Tipos de participantes*
    - Tipos de intervenciones*
    - Tipos de medidas de desenlace*
  - Métodos de búsqueda para la identificación de estudios*
  - Extracción y análisis de los datos*
- Resultados:
  - Descripción de los estudios*
  - Riesgo de sesgo de los estudios incluidos*
  - Efecto de las intervenciones*
- Discusión*
- Conclusiones de los autores:
  - Implicaciones para la práctica*
  - Implicaciones para la investigación*
- Agradecimientos
- Referencias:
  - Referencias a los estudios:
    - Estudios incluidos
    - Estudios excluidos
    - Estudios pendientes de clasificación
    - Estudios en curso.
- Otras referencias:
  - Referencias adicionales
  - Otras versiones publicadas de esta revisión
- Tablas y figuras:
  - Características de los estudios:
    - Características de los estudios incluidos (*incluye las tablas de 'Riesgo de sesgo'*)
    - Características de los estudios excluidos
    - Características de los estudios pendientes de evaluación
    - Características de los estudios en curso
- Tablas de 'Resumen de los resultados'
- Tablas adicionales
- Figuras

Información complementaria:
- Datos y análisis
- Anexos
- Comentarios:
  - Título
  - Resumen
  - Respuestas
  - Colaboradores

Sobre el artículo:
- Contribuciones de los autores
- Declaraciones de interés*
- Diferencias entre el protocolo y la revisión
- Financiación:
  - Fuentes internas
  - Fuentes externas
- Notas publicadas

2.3 Aspectos logísticos para hacer una revisión

2.3.1 Motivación para asumir una revisión

Los autores pueden estar motivados para asumir una sistemática por numerosas razones . Por ejemplo, las revisiones pueden desarrollarse en un esfuerzo por resolver una controversia en la evidencia científica, contestar preguntas cuando la práctica clínica es incierta, explorar variaciones en la práctica, confirmar la pertinencia de una práctica determinada o destacar un tema para una investigación futura. El objetivo general de las revisiones Cochrane debe ser el de resumir y ayudar a la gente a comprender la evidencia científica. Deben ayudar a la gente a tomar decisiones prácticas sobre la atención sanitaria. Este objetivo tiene implicaciones importantes en el momento de decidir si se hace o no una revisión Cochrane, en cómo formular la pregunta que guiará la revisión, cómo desarrollar los criterios de elegibilidad para guiar la inclusión de estudios basándose en la pregunta de la revisión, cómo desarrollar el protocolo y cómo presentar los resultados de la revisión.

2.3.2 Planificar el tema y el alcance de una revisión

Algunos de los puntos importantes a tener en cuenta cuando se planea una revisión y el desarrollo de su protocolo son:

- Las preguntas de la revisión deben orientar las alternativas (opciones prácticas) que afronta la gente cuando debe decidir sobre la atención sanitaria.
- Las revisiones deben abordar las variables de resultado importantes para las personas que toman decisiones sobre la atención sanitaria.
- Los autores de la revisión deben describir cómo manejarán tanto los beneficios como los efectos adversos de una intervención.
- Los métodos utilizados en una revisión deben seleccionarse para optimizar la posibilidad de que los resultados aportarán la mejor evidencia actual sobre la cual basar las decisiones, y deben ser descritos con suficiente detalle en el protocolo para que los lectores puedan comprender suficientemente los pasos planificados.
- Es importante que las personas sepan cuando la evidencia no es fiable o no hay evidencia científica sobre determinadas variables de resultado que puedan ser importantes para los que toman las decisiones. La falta de evidencia de un determinado efecto no debe confundirse con la existencia de evidencia científica sobre la ausencia de un determinado efecto.
- No es útil incluir evidencia científica con un alto riesgo de sesgo en una revisión, aún en los casos en los que la evidencia disponible no es mejor. El Capítulo 8 recoge una discusión más detallada sobre sesgo.
- De igual manera, no es útil centrarse en variables de resultado triviales sólo porque hayan sido escogidas por otros investigadores para hacer las mediciones en estudios individuales (ver Capítulo 5).
- En la medida de lo posible, es importante asumir una perspectiva internacional. La evidencia recopilada no debe estar restringida por nacionalidad o idioma sin una buena razón, la información recogida en los antecedentes sobre la prevalencia y la morbilidad debería ofrecer una visión global, y habría que hacer el esfuerzo de situar los resultados de la revisión en un contexto amplio.

2.3.3 Registrar un protocolo

El primer paso en el proceso de la revisión es acordar un tema de revisión con un Grupo Cochrane de Revisiones (CRG), los temas que cubren cada uno de los cincuenta CRG se describen en su propio alcance, publicados en la *CDSR*. Muchos CRG habrán priorizado algunas revisiones de importancia, y requerirán cumplimentar un 'documento para registro del título'. Un título será registrado posiblemente después de una discusión entre los editores

del CRG, y entonces los autores de la revisión serán invitados a presentar un protocolo. Una vez completado el protocolo se enviará al CRG para la revisión por pares de los editores y los expertos que colaboran con el CRG. Cuando éstos estén satisfechos con el protocolo (proceso que puede suponer varias discusiones) lo incluirán en el módulo del CRG para ser publicado y divulgado en el *CDSR*. Los editores y los autores no deberán incluir un protocolo en un módulo a menos que haya un firme compromiso de completar la revisión en un plazo de tiempo razonable y mantenerlo actualizado una vez esté completo.

La política de la Colaboración Cochrane es que los protocolos que no se hayan convertido en revisiones sistemáticas completas en el plazo de dos años en principio deben ser retirados de la *CDSR*. Si un protocolo es retirado por cualquier razón diferente a la de haber sido reemplazado por una revisión, la razón de su retiro deberá publicarse en la *CDSR* por un número. Posteriormente, la razón del retiro del protocolo deberá aparecer en el módulo del CRG.

2.3.4 El equipo de revisión

2.3.4.1 La importancia de un equipo

Es esencial que las revisiones Cochrane sean asumidas por más de una persona. Esto asegura que labores como la selección de estudios para elegibilidad y la extracción de datos puedan ser realizadas independientemente por al menos dos personas, aumentando así la posibilidad de detectar errores. Si más de un equipo manifiesta su interés en asumir la revisión del mismo tema, es probable que un CRG los anime a trabajar conjuntamente.

Los equipos de Revisión deben contar con expertos en el área temática que se está revisando e incluyen, o tienen acceso, a expertos en la metodología de las revisiones sistemáticas (entre los que figuran expertos en estadística). A los autores de la revisión que no tengan experiencia en el proceso de revisión sistemática se les anima a trabajar con otros que la tengan y a asistir a talleres de formación, organizados por la Colaboración (ver Sección 2.3.6). La Colaboración Cochrane está comprometida en involucrar a pacientes y usuarios de los servicios sanitarios por principios (el décimo principio de la Colaboración es permitir una amplia participación, ver Capítulo 1, Tabla 1.1a) y anima a los autores de la revisión a buscar e incorporar en el desarrollo de los protocolos y revisiones los puntos de vista de pacientes y usuarios, y las personas de varias regiones y ámbitos. Si un tema de revisión tiene una relevancia particular en una región o ámbito (por ejemplo, las revisiones sobre la malaria en los países en vías de desarrollo), se recomienda la participación de habitantes de ese lugar.

2.3.4.2 Participación de los pacientes y usuarios

La Colaboración Cochrane recomienda la participación de los usuarios de los servicios sanitarios, bien sea como parte del equipo de la revisión o en el proceso editorial. La participación del usuario ayuda a asegurar que las revisiones:

- Traten asuntos que son importantes para la gente;
- Tengan en cuenta las variables de resultado que son importantes para las personas afectadas;
- Sean accesibles para las personas que toman decisiones;
- Reflejen de manera adecuada la variabilidad de los valores y las preferencias de los individuos, y las circunstancias de la asistencia sanitaria en diferentes países.

Se sabe relativamente poco acerca de la efectividad de las diversas maneras de incluir a los usuarios en el proceso de revisión o, más ampliamente, en la investigación en el campo de la asistencia sanitaria (Nilsen 2006). Sin embargo, la Colaboración apoya por principio la participación de los usuarios. Esto está basado en nuestros principios, en el sentido común, y

la evidencia de que las opiniones y perspectivas de los usuarios a menudo difieren ampliamente de las de los profesionales sanitarios de salud y de los investigadores (Bastian 1998).

Los usuarios participan en el desarrollo de protocolos y revisiones de las siguientes maneras:

- Apoyando a los CRG para priorizar temas de revisión
- Co-autoría de revisiones
- Contribuyendo en la consulta a un usuario durante el desarrollo del protocolo y la revisión
- Realizando la revisión por pares los protocolos y las revisiones

Siempre que sean consultados usuario (o cualquier otra persona) durante el desarrollo de un protocolo o una revisión, su contribución deben ser reconocidas en la sección de Agradecimientos del protocolo o revisión.

Si la aportación a la revisión es más significativa, puede ser apropiada una inclusión formal en el listadode autores de la revisión, como debe hacerse con otros colaboradores (Ver el Capítulo 4, Sección 4.2.2).

2.3.4.3 Grupos asesores

Las revisiones sistemáticas son probablemente más relevantes para el usuario final y de mayor calidad si se cuenta con el asesoramiento de personas con un cierto grado de experiencia, tanto en el tema de interés como en la metodología (Khan 2001, Thomas 2004, Rees 2004). Dado que las prioridades de los que toman decisiones y de los usuarios pueden ser diferentes de las de los autores, es importante que los autores aborden los asuntos de importancia para los interesados e incluyan intervenciones, variables de resultado y poblaciones relevantes. Puede ser útil contar con un grupo de asesores, que incluya representación de personas con un interés en el campo (*stakeholders*), con intereses relevantes, habilidades y comprometidos. Esto puede ser de gran importancia en revisiones de las que se espera que sean de gran impacto, o en revisiones sobre intervenciones complejas relacionadas con diversos ámbitos. El recuadro 2.3.a muestra un ejemplo en el que se utilizó un grupo asesor para mejorar una revisión.

La aportación del grupo asesor deberá estar coordinado por el equipo de revisión para informar de las decisiones clave que se hayan tomado. El Effective Public Health Practice Project, en Canadá, ha encontrado que seis miembros pueden cubrir todas las áreas y es útil para las revisiones de salud pública (Effective Public Health Practice Project 2007). Sin embargo, cuanto más amplia es la revisión más experiencia requieren los miembros del grupo asesor.

Es importante considerar las necesidades de países con escasos recursos en los procesos de revisión. Para aumentar la relevancia de las revisiones sistemáticas, los autores también deben consultar a la población de los países en vías de desarrollo para identificar los temas prioritarios sobre los cuales se deben hacer las revisiones (Richards 2004). También es importante incluir población vulnerable y marginada en el grupo asesor (Steel 2001), con el fin de asegurar que las conclusiones respecto al valor de las intervenciones estén bien informadas y sean aplicables a todos los grupos de la comunidad.

Pueden desarrollarse descripciones del trabajo o perfiles personales para un grupo asesor para asegurar que las tareas requeridas están claras. Se aportan ejemplos en notas breves para los investigadores (Hanley 2000) o en la página INVOLVE (www.invo.org.uk). Los miembros del grupo asesor pueden desarrollar una o más de las siguientes tareas:

- Elaborar y redefinir decisiones acerca de las intervenciones de interés, la población que se va a incluir, variables de resultado prioritarios y, posiblemente, análisis de subgrupos;

- Aportar o sugerir materiales de apoyo importante que aclare los asuntos desde diferentes perspectivas;
- Ayudar a interpretar los resultados de la revisión;
- Diseñar un plan de divulgación y contribuir en la divulgación entre grupos relevantes.

Recuadro 2.3.a: Un ejemplo de los beneficios de utilizar un grupo asesor en el proceso de planificación

Una revisión de prevención del VIH en hombres que mantienen relaciones sexuales con hombres (Rees 2004) empleó métodos de consenso explícitos para diseñar la revisión con la ayuda de médicos, delegados e investigadores. Se convocó un grupo asesor con miembros de organizaciones académicas, de investigación, políticas y de servicios, así como con representantes de organizaciones benéficas, y organizaciones que habían surgido y hablaban en nombre de personas que convivían o estaban afectadas de VIH y sida. El grupo se reunió tres veces durante el curso de la revisión.

Se entregó al grupo información previa sobre la revisión propuesta: su alcance, bases conceptuales, objetivos, preguntas de investigación, etapas y métodos. La discusión se centró en la relevancia política y en el antecedente/contexto político de la revisión; los criterios de elegibilidad de los estudios (intervenciones, variables de resultado, subgrupos de participantes); estrategias de divulgación; y cronograma. Dos votaciones identificaron y priorizaron los variables de resultado de interés para su análisis. En discusiones abiertas se identificaron subgrupos de hombres vulnerables. Mediante las discusiones del grupo asesor se precisó un marco para definir las intervenciones de interés.

La revisión siguió esta guía al adoptar las intervenciones identificadas, las poblaciones y las variables de resultado para precisar los criterios de inclusión, realizar un metanálisis y un análisis de subgrupo. El producto resultante incluyó evidencia sintetizada relacionada directamente con desigualdades en salud.

2.3.5 Software Cochrane para autores de revisión y apoyo editorial a los Grupos Cochrane de Revisión

Para apoyar la preparación y la supervisión editorial de las revisiones Cochrane, la Colaboración Cochrane utiliza el Cochrane *Information Management System* (IMS). El IMS tiene dos componentes principales: un software para la redacción de las revisiones, Review Manager (RevMan), y un servidor central para el manejo de documentos y los datos de contacto, Archie. El IMS funciona como la infraestructura electrónica de la Colaboración Cochrane y facilita una colaboración eficiente entre los profesionales que trabajan en el seno de los CRGs y sus equipos de autores, que a menudo trabajan en diferentes continentes.

El RevMan es una herramienta de uso obligatorio para los autores Cochrane que utilizan en la preparación y el mantenimiento de los protocolos y revisiones en el formato descrito en la Sección 2.2. El software se desarrolla mediante un proceso continuo de consulta con sus usuarios y con los expertos en metodología Cochrane para apoyar las directrices y las guías de las revisiones Cochrane y así aportar métodos de análisis mejorados, ayuda 'en línea' y mecanismos para la detección de errores.

Además de apoyar la preparación de una revisión Cochrane de Intervención, el RevMan contribuye en la preparación de revisiones Cochrane sobre Metodología , revisiones Cochrane sobre el rendimiento de pruebas Diagnósticas, y Revisiones globales (*overviews*) de revisiones (ver Capítulo 22).

El uso de RevMan es gratis para los autores que preparan una revisión Cochrane y para instituciones académicas. Las empresas comerciales pueden usar el software si compran una licencia. Se brinda apoyo técnico sólo a los autores que hayan registrado sus revisiones Cochrane en el seno de un CRG.

Mientras que el RevMan se usa para preparar y editar revisiones, Archie se utiliza para guardar diferentes borradores y versiones publicadas de las revisiones. Al reunir todas las versiones relevantes de una revisión centralmente, el sistema facilita el acceso a la última versión publicada de una revisión cuando ésta se requiere para una actualización. A través de Archie los autores también pueden ver las versiones previas de una revisión y comparar dos versiones de la misma revisión para identificar los cambios introducidos de una versión a la siguiente. Además, los autores mantienen sus detalles de contacto y pueden acceder a los detalles de contacto de sus coautores y de su equipo editorial. Los autores de una revisión Cochrane pueden tener acceso a Archie contactando con la base editorial de su CRG.

El IMS se desarrolla y mantiene en el Centro Cochrane Nórdico. El desarrollo continuo del IMS es supervisado por el Grupo Cochrane *Information Management System* con el apoyo de los grupos asesores correspondientes. Se dispone de más información acerca del software de la Colaboración Cochrane, así como las últimas versiones y los cambios planificados, en la página del IMS: www.cc-ims.net.

2.3.6 Formación

Es importante asegurar que las personas que contribuyen en el trabajo de la Colaboración tengan el conocimiento, las habilidades y el apoyo que necesitan para hacer un buen trabajo. Esta formación pueden necesitarla los autores de una revisión, los editores, los revisores por pares, los coordinadores de CRG y Coordinadores de Búsqueda de Ensayos, los buscadores manuales, formadores y usuarios de revisiones Cochrane. Aquí nos vamos a centrar en las necesidades de formación de los autores de revisión y los editores, para ayudarles a preparar y mantener revisiones de alta calidad.

Mientras algunos autores de revisión que se unen a un CRG tienen formación y experiencia en realizar una revisión sistemática, otros no la tienen. Además de los materiales de formación y el apoyo a los autores que ofrecen varios CRG, los Centros Cochrane tienen la responsabilidad de trabajar con Grupos de Métodos para desarrollar materiales de formación basados en el Manual y organizar talleres de formación para miembros de los CRG. Cada CRG tiene la responsabilidad de asegurar que los autores de la revisión tienen la formación y el apoyo metodológico adecuados. Los materiales para la formación y las oportunidades para obtenerla se elaboran y actualizan continuamente para reflejar el desarrollo necesario de la Colaboración así como sus directrices y sus guías metodológicas.

La formación para los autores de revisión se lleva a cabo en varios países por parte de los Centros Cochrane, los Grupos de Métodos y los CRG. La programación de formación aparece en la página de formación de la Colaboración Cochrane (training.cochrane.org/), junto con otras fuentes de formación como los Open Learning Material de la Colaboración Cochrane. Los detalles sobre los Centros Cochrane se encuentran en www.cochrane.org.

2.3.7 Procedimientos editoriales de un Grupo Cochrane de Revisión

El equipo editorial del CRG es en último término el responsable de la decisión de publicar una revisión Cochrane en su módulo. Esta decisión se tomará a tras la revisión por pares y modificaciones pertinentes por parte de los autores de la revisión. Esto puede tomar varias rondas de edición.

El equipo editorial de cada CRG es el responsable de mantener un módulo, que contiene la información acerca del Grupo, incluidos los procesos editoriales. Cualquiera de los métodos específicos utilizados por el CRG más allá de los especificados en el Manual deben ser documentados en su módulo, incluidos:

- Los métodos utilizados para revisar los protocolos;
- Los criterios de elegibilidad para considerar la inclusión de estudios en las revisiones;
- Los métodos de búsqueda y estrategias específicas de búsqueda diseñados para desarrollar y mantener el Registro Especializado utilizado por el CRG, y el método para ofrecer artículos potencialmente relevantes o las publicaciones a texto completo a los autores;
- Métodos adicionales de búsqueda, sobre los que se instruye a los autores para utilizarlos rutinariamente;
- Los métodos habituales utilizados para seleccionar estudios para las revisiones y algunos formularios para evaluar la inclusión de estudios;
- Los criterios o métodos diferentes fuera del alcance de la tabla de 'Riesgo de sesgos' utilizada para evaluar los estudios incluidos;
- Los métodos habituales empleados para la extracción de datos y plantillas para la extracción de datos.

The Cochrane Library publica las descripciones de los métodos adicionales específicos utilizados por cada CRGcomo parte de su módulo. Los autores deben familiarizarse con los contenidos del módulo de su Grupo.

2.3.8 Recursos para una revisión sistemática

Las revisiones Cochrane son preparadas por autores que trabajan en el seno de un CRG. Cada CRG tiene un equipo editorial responsable de producir un módulo de las revisiones editadas para su difusión a través de la *CDSR* en *The Cochrane Library.*

Como la Colaboración Cochrane está organizada alrededor de los CRG, es importante que cada autor esté vinculado con uno de estos Grupos desde el comienzo del proceso. Además de asegurar que las revisiones Cochrane están adecuadamente realizadas, esta estructura reduce el peso que recae individualmente sobre los autores ya que los equipos editoriales son los responsables de aportar la mayoría o todo el apoyo en:

- Dirigir las búsquedas sistemáticas de los estudios relevantes y coordinar la distribución de los estudios potencialmente relevantes entre los autores;
- Establecer las normas y los procedimientos específicos para el CRG;
- Asegurarse de que los autores reciban el apoyo metodológico que necesitan.

El principal recurso que los autores necesitan es su propio tiempo. La mayoría de los autores aportarán su tiempo sin ninguna remuneración económica porque lo verán como un esfuerzo personal por mantenerse actualizados en su área de interés. En algunos casos los autores pueden necesitar recursos adicionales, o al menos poderle justificar la cantidad de tiempo requerido para una revisión sistemática a colegas que aún no entienden qué implica una revisión sistemática o su importancia.

La cantidad de tiempo requerido puede variar de acuerdo con el tema de la revisión, el número de estudios, los métodos utilizados (por ejemplo, los esfuerzos para obtener información no publicada), la experiencia de los autores, y los tipos de apoyo aportados por el equipo editorial. La carga de trabajo asociada con asumir una revisión es por lo tanto muy variable. Sin embargo, la consideración de las tareas correspondientes y el tiempo requerido para cada una de ellas puede ayudar a los autores a calcular la cantidad de tiempo que necesitarán. Estas tareas incluyen formación, reuniones, desarrollo del protocolo, búsqueda de estudios, evaluación de citas y publicaciones de estudios para decidir su elegibilidad, evaluando el riesgo de sesgo de los estudios incluidos, extrayendo datos, buscando datos no

disponibles y estudios no publicados, analizando los datos, interpretando los resultados y escribiendo la revisión, manteniendo la revisión actualizada.

Un cronograma con fechas prioritarias para cumplir tareas clave puede ayudar a programar el tiempo necesario para completar una revisión. Tales objetivos pueden variar de una revisión a otra. Los autores, junto con el equipo editorial del CRG, deben determinar un marco de tiempo apropiado para una revisión específica. Un ejemplo de un calendario con fechas elegidas se encuentra en el recuadro 2.3.b.

Los recursos que pueden necesitarse para estas tareas, además del tiempo de los autores, incluyen:

- Búsqueda (la identificación de los estudios es básicamente responsabilidad del equipo editorial del CRG. Sin embargo, los autores pueden compartir esa responsabilidad y puede ser conveniente buscar bases de datos adicionales para una revisión específica);
- Ayuda de trabajo bibliotecario, préstamos entre bibliotecas y fotocopias;
- Un segundo autor para evaluar los estudios a incluir, evaluar el riesgo de sesgo de los estudios incluidos, obtener datos y revisar la extracción de datos y los análisis;
- Apoyo estadístico para sintetizar (si es el caso) los resultados de los estudios incluidos;
- Equipo (por ejemplo, *hardware* y *software* informáticos);
- Suministros y servicios (costes de llamadas teléfonicas de larga distancia, conexión a Internet, faxes, papel, impresora, fotocopiadora, suministros para ordenadores y audiovisuales);
- Espacio en el despacho para uso del equipo de apoyo;
- Fondos para viajes.

Recuadro 2.3.b: Calendario para una revisión Cochrane

Mes	Actividad
1–2	Preparación del protocolo.
3–8	Búsquedas de estudios publicados y no publicados
2–3	Prueba piloto de criterios de elegibilidad.
3–8	Evaluación de la inclusión de estudios.
3	Prueba piloto de evaluación de ‘Riesgo de sesgo’.
3–10	Evaluaciones de la validez.
3	Prueba piloto de la extracción de datos.
3–10	Extracción de datos.
3–10	Entrada de datos en RevMan.
5–11	Seguimiento de datos perdidos.
8–10	Análisis.
1–11	Preparación del informe de la revisión.
12	Mantenimiento de la revisión actualizada.

2.3.9 Búsqueda de financiación

Muchas organizaciones frecuentemente aportan financiación para revisiones sistemáticas de temas prioritarios. Aquí se incluyen agencias para la financiación de investigación, organizaciones que aportan o financian servicios de salud, los responsables de la evaluación de tecnologías sanitarias, y aquellos que están involucrados en el desarrollo de guías para la práctica clínica.

La Colaboración tiene la política de que ni la preparación de las revisiones Cochrane ni los costes de la infraestructura para un CRG pueden financiarse a través de una fuente o agencia comercial que tenga un interés creado en la revisión(ver Sección 2.6).

2.4 Publicación de revisiones Cochrane en revistas biomédicas y libros

Los autores de una revisión pueden estar interesados en buscar publicaciones conjuntas de las revisiones Cochrane en otras revistas sanitarias revisadas por pares, especialmente en aquellas revistas que han mostrado interés por hacer publicaciones conjuntas de las revisiones Cochrane. Para la Colaboración Cochrane sólo hay una condición fundamental para la publicación conjunta: las revisiones Cochrane deben permanecer libres para ser difundidas en cualquier medio, sin restricción de ninguno de ellos. Para asegurar esto, los autores de Cochrane conceden a la Colaboración licencias en todo el mundo para estas actividades, y no firman derechos exclusivos con ninguna revista o editor. Una revista puede pedir un derecho no exclusivo que le permita publicar y volver a publicar una revisión, pero esto no le impide a la Colaboración Cochrane publicar la revisión en cualquier forma que ella considere apropiada. Para volver a publicar material publicado en la *CDSR* en cualquier parte, especialmente en revistas científicas impresas, los autores deben cumplimentar un formulario de 'permiso para publicar' disponible en el Manual Cochrane (www.cochrane.org/admin/manual.htm), junto con una explicación de los procedimientos a seguir.

Se recomienda encarecidamente que los autores no publiquen las revisiones Cochrane en revistas antes de que estén listas para ser publicadas en la *CDSR*. Esto se aplica especialmente a los directores de los Centros y a editores de los CRG. Sin embargo, a veces las revistas insisten en que la publicación de las revisiones en la *CDSR* no debe preceder a la publicación impresa. Cuando esto sucede, los autores deben acceder a enviar una revisión para su publicación en la revista después de un acuerdo con el editor de su CRG y antes de su publicación en *CDSR*. La publicación en la revista no debe estar sujeta a tiempos prolongados de producción, y los autores no deben retardar demasiado la publicación de una revisión Cochrane ni por demoras en una revista o con el fin de reenviar su revisión a otra revista.

Las revistas también pueden pedir modificación de una revisión por razones editoriales o de contenido. La revisión externa por pares aportada por las revistas puede resaltar el valor de la revisión y deben ser bienvenidas. Las revistas generalmente requieren revisiones más cortas que las publicadas en la *CDSR*. La reducción selectiva de las revisiones puede ser apropiada, pero no debe haber diferencias sustanciales entre la revisión tal y como se publica en la revista y la de la *CDSR*. Si una revisión se publica en una revista debe anotarse que una versión más completa y actualizada de la revisión se encuentra disponible en la *CDSR*. Usualmente esto debe hacerse incluyendo en la introducción una declaración como esta: 'Una revisión más detallada será publicada y actualizada en la *Base de datos Cochrane para Revisiones sistemáticas'*. La referencia debe remitir al protocolo de la revisión publicado en la *CDSR*. Una declaración similar debe incluirse en la introducción si una revisión es publicada en la *CDSR* antes de que la versión de la revisión sea publicada en una revista. Después de que una revisión Cochrane ha sido publicada en una revista, se debe publicar una referencia a esta publicación en la revista bajo el encabezamiento 'Otras versiones publicadas de esta revisión'. También se sugiere a los autores agregar la siguiente declaración a las versiones de revisiones Cochrane que son publicadas en las revistas:

'Este artículo está basado en una revisión Cochrane publicada previamente [o recientemente corregida de manera sustancial, según corresponda] en The Cochrane Library YYYYY, Volumen X (para más información ver http://www.thecochranelibrary.com/). Las revisiones Cochrane son actualizadas periódicamente a medida que aparece nueva evidencia y en respuesta a los comentarios, y debe consultarse The Cochrane Library para las versiones más recientes de la revisión'.

Se debe añadir una modificación del descargo de responsabilidad (*disclaimer*) en las revisiones Cochrane publicadas en revistas:

´Los resultados de una revisión Cochrane pueden ser interpretados de manera diferente, dependiendo de las perspectivas y circunstancias de las personas. Por favor analice cuidadosamente las conclusiones presentadas. Son las opiniones de los autores de la revisión y no son necesariamente compartidas por la Colaboración Cochrane.'

El párrafo anterior debe ser enviado a los editores de la revista a la cual se presenta la revisión para su publicación, y se debe enviar copia de la carta de presentación al equipo editorial del CRG para su información. Esta política y este procedimiento podrían ser nuevos para algunos editores de revistas y puede necesitar discusión directa con ellos. El grupo editorial del CRG debe ser informado de cualquiera de los problemas encontrados en este proceso. Se sugiere que el siguiente párrafo sea incluido en las cartas de presentación a los editores de revistas:

'Esta revisión sistemática ha sido preparada bajo el auspicio de la Colaboración Cochrane, una organización internacional que se propone ayudar a la gente a tomar decisiones bien informadas sobre la atención sanitaria, preparando, manteniendo y promocionando el acceso a las revisiones sistemáticas sobre los efectos de las intervenciones en salud. La política de publicación de la Colaboración le permite a las revistas publicar revisiones, con prioridad si se requiere, pero también le permite a la Colaboración Cochrane publicar y divulgar tales revisiones. Las revisiones Cochrane no pueden estar sujetas a los derechos exclusivos pedidos por algunas revistas'.

2.5 Publicación de revisiones publicadas previamente como revisiones Cochrane

La mayoría de las revisiones que han sido realizadas por autores fuera de la Colaboración Cochrane (denominados aquí 'revisiones publicadas previamente') necesitan un trabajo adicional importante antes de poder ser publicadas como una revisión Cochrane en la *CDSR*. En vista de este trabajo adicional y de las diferencias sustanciales con la revisión previamente publicada, la revisión Cochrane puede considerarse como una nueva publicación. La versión publicada previamente de la revisión debe ser referenciada en la revisión Cochrane bajo el encabezamiento 'Otras versiones publicadas de esta revisión'. Sin embargo, generalmente no es necesario pedir la autorización de los editores de la revisión publicada previamente.

Ocasionalmente una revisión Cochrane puede ser tan similar a la revisión publicada previamente que el único cambio sea el formato de la revisión. En estos casos los autores deben obtener permiso de los editores de la revisión publicada previamente antes de publicar la revisión en la *CDSR*. Si los autores tienen dudas sobre la solicitud del permiso se les debe recomendar que lo pidan. Esto no debe representar un problema ya que se presenta mucho antes de la presentación prevista a la *CDSR*. Si se sabe con anticipación de que hay interés de publicar en la *CDSR* la versión de una revisión ya publicada en una revista, los autores no deben cederle derechos de exclusividad a la revista (ver Sección 2.4). La Colaboración Cochrane no requiere derechos de autor exclusivos. No es por lo tanto un problema publicar la versión de una revisión Cochrane en una revista después de que esta haya sido publicada en la *CDSR*, siempre que no se le llame revisión Cochrane y que se reconozca que está basada en una revisión Cochrane (ver Sección 2.4)

2.6 Declaración de interés y patrocinio comercial

Las revisiones Cochrane deben estar libres de cualquier sesgo real o aparente debido al recibo de cualquier beneficio en efectivo o en especie, cualquier atención, o cualquier subsidio derivado de cualquier fuente que pueda tener o ser interpretado como susceptible de tener interés en el resultado de la revisión. Debe existir una clara barrera entre la producción de las revisiones Cochrane y cualquier financiación de fuentes comerciales con intereses financieros en las conclusiones de las revisiones Cochrane. Así, el patrocinio de una revisión Cochrane por cualquier fuente o fuentes comerciales (como se definió arriba) está prohibido. Se permiten otros patrocinadores, pero a un patrocinador no le está permitido retrasar o impedir la publicación de una revisión Cochrane, y un patrocinador no debe tener la capacidad de interferir con la independencia de los autores de las revisiones respecto a la conducción de sus revisiones. El protocolo de una revisión Cochrane debería mencionar específicamente que un patrocinador no puede impedir ciertas variables de resultado sean evaluadas en la revisión.

Estas reglas también se aplican a 'productos derivados' (que contienen revisiones Cochrane), así que los patrocinadores comerciales no pueden impedir o influenciar lo que puede ser incluido en tales productos. El recibo de beneficios de cualquier fuente de patrocinadores de investigación debe ser reconocido y los conflictos de interés deben ser revelados en la *CDSR* y en otras publicaciones que pertenezcan a la Colaboración.

El código de conducta de la Colaboración Cochrane para evitar potenciales conflictos de interés financiero aparece en el Recuadro 2.6.a. Si una propuesta para asumir una revisión origina la duda de serios conflictos de interés, debe ser enviada al árbitro de patrocinio de la Colaboración (fundingarbiter@cochrane.org) para su revisión. No es obligatorio enviar propuestas de patrocinio al Centro Cochrane local o al *Steering Group* antes de aceptarlas. Sin embargo, esto sería deseable en los casos de donaciones restringidas o en cualquier donación que parezca entrar en conflicto con los principios generales anotados anteriormente.

Es imposible abolir el conflicto de interés, ya que la única persona que no tiene algún interés creado en un tema es alguien que no conoce nada acerca de él (Smith 1994). Los conflictos financieros de interés, que son la causa de la mayoría de ellos, pueden y deben ser evitados, pero deben ser declarados si los hay.

Cualquier interés secundario (tales como conflictos personales) que pueda influenciar demasiado los juicios hechos en una revisión (en relación, por ejemplo, con la inclusión o exclusión de estudios, evaluaciones del riesgo de sesgo en los estudios incluidos o en la interpretación de los resultados) debe ser declarado. Un ejemplo corriente ocurre cuando el autor de una revisión también es autor de un estudio potencialmente elegible. Esto debe ser declarado en la revisión y, cuando sea posible, debe hacerse una evaluación independiente de elegibilidad y riesgo de sesgo por un segundo autor que no tenga conflicto de intereses.

Descubrir un conflicto de intereses no reduce necesariamente el valor de una revisión y no significa deshonestidad. Sin embargo, los conflictos de interés pueden influenciar los juicios de formas sutiles. Los autores deben permitir que los editores de su Grupo Cochrane de Revisión conozcan los potenciales conflictos aun cuando ellos confíen en que sus juicios no son o no serán influenciados. Los editores pueden decidir que no se justifica revelarlos o pueden decidir que los lectores deben conocer ese conflicto de intereses para que sean ellos mismos los que decidan cuán importante es. La decisión de publicar o no tal información deben hacerla conjuntamente autores y editores.

Para ayudar a asegurar la integridad real y la integridad aparente de las revisiones Cochrane, todos los autores deben firmar las evaluaciones relevantes en el formulario, dándole a la Colaboración Cochrane permiso para publicar su revisión además de las declaraciones de

interés, y el equipo editorial de cada CRG debe también revelar cualquier conflicto potencial de interés que pueda tener, tanto en su módulo como dentro de las revisiones relevantes.

Recuadro 2.6.a: Código de Conducta de la Colaboración Cochrane para evitar posibles conflictos de interés financiero

Principio General
La actividad fundamental de la Colaboración Cochrane es coordinar la preparación y el mantenimiento de las revisiones sistemáticas sobre los efectos de las intervenciones sanitarias, realizadas por autores individuales de acuerdo con los procedimientos establecidos por la Colaboración Cochrane. La realización de la revisión debe estar libre de cualquier sesgo real o aparente, causado por recibir cualquier beneficio en dinero o en especie, cualquier atención o cualquier subsidio derivado de cualquier fuente que pueda tener, o ser susceptible de tener, un interés en el resultado de la revisión. Todas las entidades que constituyen la Colaboración Cochrane deben aceptar este Principio General como condición para participar en la Colaboración.

Política
(i) Debe reconocerse la recepción de beneficios de cualquier fuente patrocinadora de investigación y los conflictos de interés deben informarse a la Base de Datos Cochrane de Revisiones Sistemáticas y a otras publicaciones que pertenezcan a la Colaboración Cochrane.

(ii) Si un autor participa en un estudio incluido en su revisión debe reconocerlo, ya que esto puede interpretarse como un posible conflicto de interés.

(iii) Si un proyecto plantea la duda de un conflicto de interés serio, éste debe dirigirse al Centro Cochrane local para su revisión (y el *Steering Group* debe ser informado). Si el asunto involucra un Centro Cochrane, el asunto debe ser remitido al Steering Group.

(iv) No es obligatorio enviar las propuestas de financiación al Centro Cochrane local o al *Steering Group* antes de aceptarlas. Sin embargo, la revisión de tales propuestas serían deseables en caso de donaciones restringidas o en cualquier donación que aparezca en conflicto con el Principio General.

(v) El *Steering Group* debe recibir (y revisar al menos anualmente) la información de todos los fondos externos aceptados por entidades Cochrane. El *Steering Group* utilizará esta información para preparar y distribuir un informe anual de los posibles conflictos de interés analizados a solicitud de la Colaboración Cochrane y del uso de recursos externos.

(vi) El *Steering Group* está considerando la formación de un Sub-Grupo de Ética para analizar los posibles conflictos de interés, ofrecer recomendaciones para su solución y considerar sanciones apropiadas para reparar infracciones del Principio General.

2.7 Información del Capítulo

Editores: Sally Green y Julian PT Higgins.

La versión en inglés de este capítulo debe citarse como: Green S, Higgins JPT (editors). Chapter 2: Preparing a Cochrane review. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Autores participantes (desde marzo 2005): Ginny Brunton, Sally Green, Julian Higgins, Monica Kjeldstrøm, Nicki Jackson, Sandy Oliver.

Agradecimientos: Esta sección se ha construido sobre versiones previas del *Manual*. Para detalles de los autores y de los editores anteriores del Manual, ver el Capítulo 1 (Sección 1.4). Agradecemos a Chris Cates, Carol Lefebvre, Philippa Middleton, Denise O'Connor y Lesley Stewart por sus comentarios sobre los borradores desde marzo 2005.

2.8 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.

Effective Public Health Practice Project 2007
Effective Public Health Practice Project. Effective Public Health Practice Project [Updated 25 October 2007]. Available from: http://www.city.hamilton.on.ca/PHCS/EPHPP (accessed 1 January 2008).

Hanley 2000
Hanley B, Bradburn J, Gorin S, Barnes M, Goodare H, Kelson M, Kent A, Oliver S, Wallcraft J. Involving Consumers in Research and Development in the NHS: briefing notes for researchers. Winchester (UK): Help for Health Trust, 2000. Available from www.hfht.org/ConsumersinNHSResearch/pdf/involving_consumers_in_rd.pdf.

Khan 2001
Khan KS, ter Riet G, Glanville J, Sowden AJ, Kleijnen J (editors). Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for those Carrying Out or Commissioning Reviews (CRD Report Number 4) (2nd edition). York (UK): NHS Centre for Reviews and Dissemination, University of York, 2001.

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

Nilsen 2006
Nilsen ES, Myrhaug HT, Johansen M, Oliver S, Oxman AD. Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material. Cochrane Database of Systematic Reviews 2006, Issue 3. Art No: CD004563.

Rees 2004
Rees R, Kavanagh J, Burchett H, Shepherd J, Brunton G, Harden A, Thomas S, Oakley A. HIV Health Promotion and Men who have Sex with Men (MSM): A systematic review of research relevant to the development and implementation of effective and appropriate interventions. London (UK): EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, 2004.

Richards 2004
Richards T. Poor countries lack relevant health information, says Cochrane editor. BMJ. 2004;328:310.

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

Steel 2001
Steel R. Involving marginalised and vulnerable groups in research: a discussion document. Consumers in NHS research [2001]. Available from: http://www.invo.org.uk/pdf/Involving_Marginalised_Groups_in_Research.pdf (accessed 1 January 2008).

Thomas 2004
Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. Worldviews on Evidence-Based Nursing. 2004;1:165-184.

ÍNDICE
CAPÍTULO 3: MANTENIMIENTO DE LAS REVISIONES: ACTUALIZACIONES, ENMIENDAS Y COMENTARIOS

PUNTOS CLAVE

3.1 INTRODUCCIÓN

3.1.1 ¿Por qué mantener una revisión?
3.1.2 ¿Con qué frecuencia debe ser revisada una revisión?

3.2 ALGUNAS DEFINICIONES IMPORTANTES

3.2.1 Introducción
3.2.2 Actualizaciones y enmiendas
3.2.3 Cita de las versiones de revisiones y protocolos Cochrane
Recuadro 3.2.a: Eventos que dan lugar a una Cita de la versión de un protocolo o revisión Cochrane
Figura 3.2.a: Resumen de cambios en las revisiones Cochrane
Figura 3.2.b: Resumen de cambios en los protocolos Cochrane

3.2.4 APLICACIÓN DE TÉRMINOS A LOS PROTOCOLOS COCHRANE

3.2.4.1 Enmiendas a los protocolos
3.2.4.2 Nueva cita de la versión del protocolo
Recuadro 3.2.b: Criterios para crear una nueva cita de la versión de un protocolo Cochrane
3.2.4.3 Ejemplos de cambios en los protocolos que no implican una nueva cita de la versión

3.2.5 APLICACIÓN DE TÉRMINOS A LAS REVISIONES COCHRANE

3.2.5.1 Actualizaciones de las revisiones
3.2.5.2 Enmiendas a las revisiones
3.2.5.3 Nuevas citas de la versión de las revisiones
Recuadro 3.2.c: Criterios para crear una nueva cita de la versión de una revisión Cochrane
3.2.5.4 Ejemplos de cambios en las revisiones que no implican una nueva cita de la versión

3.3 FECHAS IMPORTANTES ASOCIADAS CON LAS REVISIONES COCHRANE

3.3.1 Introducción
3.3.2 Fecha en que la revisión fue evaluada como actualizada
Recuadro 3.3.a: Guía para declarar actualizada una revisión
3.3.3 Fecha de la búsqueda
3.3.4 Fecha esperada de la próxima etapa
3.3.5 Fecha de la última corrección
3.3.6 Fecha en que se declaró que la revisión no necesita más actualizaciones

3.4 CONSIDERACIONES CUANDO SE ACTUALIZA UNA REVISIÓN COCHRANE

3.4.1 Por dónde comenzar

3.4.2 Actualización de una revisión con una pregunta de revisión sin cambios
3.4.2.1 Ejecución de una nueva búsqueda
3.4.2.2 Actualización de las revisiones cuando no se encuentran nuevos estudios
3.4.2.3 Actualización de revisiones cuando se encuentran nuevos estudios
3.4.3 Cambios en la pregunta de la revisión y los criterios de selección
3.4.4 Partición de Revisiones
3.4.5 Actualización de la metodología de una revisión
3.4.6 Otros cambios en la revisión
3.4.7 Proceso editorial

3.5 TABLAS 'NOVEDADES' E HISTORIAL

3.5.1 Los eventos en 'Novedades'
3.5.2 Completando la tabla 'Qué hay de nuevo'
Tabla 3.5.a: Eventos posibles para la tabla 'Novedades' de los protocolos
Tabla 3.5.b: Eventos posibles para la tabla 'Novedades' de las revisiones
3.5.3 Tabla de Historial

3.6 Incorporar y orientar la retroalimentación o comentarios a una revisión Cochrane

3.7 INFORMACIÓN DEL CAPÍTULO

3.8 REFERENCIAS

CAPÍTULO 3: MANTENIMIENTO DE LAS REVISIONES: ACTUALIZACIONES, ENMIENDAS Y COMENTARIOS

Autores: Julian PT Higgins, Sally Green y Rob JPM Scholten.

Puntos clave

- Las revisiones sistemáticas que no se actualizan pueden quedar desfasadas o llegar a ser erróneas;
- La política de la Colaboración Cochrane es que las revisiones Cochrane de Intervenciones se deben actualizar a los dos años, o incluir un comentario que explique por qué no se ha hecho;
- Cualquier cambio en una revisión Cochrane es o una actualización o una enmienda. Las actualizaciones implican la búsqueda de nuevos estudios, cualquier otro cambio es una enmienda;
- Las revisiones Cochrane son citadas por su versión. Este Capítulo incluye una lista de criterios para determinar cuándo resulta apropiado citar una nueva versión;
- Además de la búsqueda de nuevos estudios, la actualización de una revisión Cochrane puede incluir la reformulación de la pregunta de revisión y la incorporación de nuevos métodos;
- Los comentarios realizados a las revisiones Cochrane contribuyen al proceso de actualización y mantenimiento;
- La 'Fecha en que la revisión fue evaluada como actualizada' es anotada por los autores y publicada al comienzo de la revisión. Los criterios para evaluar una revisión como actualizada se establecen en este Capítulo.

3.1 Introducción

3.1.1 ¿Por qué mantener una revisión?

El principal objetivo de una revisión Cochrane es aportar la mejor evidencia disponible y más actualizada sobre los efectos de las intervenciones para que pueda ser utilizada por los usuarios, los clínicos y los responsables de las políticas de salud para que puedan tomar decisiones sanitarias bien informadas.

Como la evidencia sobre un tema dado es generalmente dinámica y evoluciona continuamente, el incorporar estudios adicionales en la medida en que están disponibles puede cambiar los resultados de una revisión sistemática (Chalmers 1994). Por lo tanto, las revisiones sistemáticas que no reciben un mantenimiento corren el riesgo de quedar desfasadas e incluso de llegar a ser erróneas. Un aspecto importante de las revisiones Cochrane es que los autores de las revisiones están comprometidos, no sólo a preparar revisiones sistemáticas de la evidencia, sino también a mantenerlas (y actualizarlas) de manera periódica.

3.1.2 ¿Con qué frecuencia debe ser revisada una revisión?

Actualmente, hay poca evidencia empírica disponible que permita tomar decisiones informadas acerca de cuál sería un enfoque razonable y eficiente para revisar la evidencia de las revisiones Cochrane, aunque existen algunas guías (Moher 2007, Shojania 2007a, Shojania 2007b). La política de la Colaboración Cochrane es que las revisiones deben actualizarse a los dos años, o incluir un comentario que explique por qué no se ha hecho. Definimos el término ‘actualizar’ en la Sección 3.2.2. El período de dos años empieza en la misma fecha en la que la revisión fue calificada como actualizada (ver Sección 3.3.2).

Además de la posible disponibilidad de nueva evidencia, hay otros aspectos que pueden apoyar la necesidad de revisar la revisión. Por ejemplo, en el campo clínico pueden haberse desarrollado nuevas herramientas o marcadores para caracterizar subgrupos, pueden estar disponibles nuevos tratamientos, o pueden estar en uso nuevas medidas de desenlace (o métodos de medida más refinados para desenlaces existentes). Además, los avances en los métodos para realizar revisiones Cochrane pueden llevar a la necesidad de revisar una revisión.

Mientras realizan la revisión, los autores pueden estar en condiciones de juzgar si se está produciendo investigación relevante de manera frecuente, y por lo tanto están en capacidad de predecir y sugerir la necesidad de actualizar la revisión con más frecuencia. Por el contrario, en otros campos de la salud, los datos aparecen muy lentamente o probablemente no aparecerán, y una revisión preparada hace muchos años puede resultar aún vigente y valiosa. En estos casos, actualizar una revisión cada dos años puede ser innecesario y un desperdicio (Chapman 2002). Se les aconseja a los autores de la revisión discutir con su Grupo Cochrane de Revisión (CRG sigla en inglés) la idea de que su revisión no necesita ser actualizada al menos cada dos años. La razón por la que la revisión no se actualiza de acuerdo con la política de la Colaboración debe declararse en la sección de la revisión ‘Notas publicadas’.

3.2 Algunas definiciones importantes

3.2.1 Introducción

Introducimos y explicamos aquí algunas definiciones importantes utilizadas por la Colaboración Cochrane, relacionadas con el mantenimiento de las revisiones y su aplicación en la publicación de las revisiones. La Sección 3.3 tiene que ver específicamente con las definiciones y el uso de las fechas que describen eventos asociados con la revisión. Aunque mucha de esta información detallada es técnica, los autores necesitarán una comprensión de estos asuntos para asegurarse del uso correcto de los términos y fechas en su revisión y cuando deban llenar los campos relacionados en RevMan.

3.2.2 Actualizaciones y enmiendas

Cualquier cambio en una revisión Cochrane es o una actualización o una enmienda.

Una actualización *debe* incluir una búsqueda de estudios nuevos. Si se encuentran estudios nuevos, deben agregarse a la sección correspondiente de la revisión como estudios incluidos, excluidos o en marcha (o 'Estudios pendientes de clasificación', si han fallado todos los esfuerzos razonables para clasificarlos en una de estas categorías), antes de calificar la revisión revisada como una actualización (ver Sección 3.2.5.1).

Cualquier otro cambio en una revisión Cochrane, y cualquier cambio en un protocolo, es una enmienda, pudiendo implica una mayor o menor cantidad de trabajo. Estos términos, y cuándo aplicarlos, se describirán con mayor detalle en la Sección 3.2.3.

3.2.3 Cita de las versiones de revisiones y protocolos Cochrane

Cada publicación de una revisión o de un protocolo Cochrane tiene una cita de la versión vigente. En el caso de las *revisiones*, la cita de la versión vigente se considera una nueva publicación principal y genera un nuevo registro en las bases de datos de referencias bibliográficas tales como MEDLINE o ISI. Los *protocolos* no tienen una cita en MEDLINE o ISI. Los eventos que conducen a la creación de una cita de la nueva versión se encuentran en el Recuadro 3.2.a.

Algunas revisiones implican cambios importantes (actualizaciones o correcciones) que generan una nueva cita en la *Base de Datos Cochrane de Revisiones Sistemáticas (CDSR sigla en inglés)* y un nuevo registro en MEDLINE e ISI (por ejemplo, cambios en las conclusiones, autores o corrección de errores importantes). A esto lo denominamos nueva cita de la versión vigente. Además, algunas nuevas citas de la versión merecen ser destacadas adicionalmente en la *CDSR* (por ejemplo, usando una etiqueta) -en particular las que cambian sus conclusiones, por lo que deben ser leídas de nuevo-. Nos referimos a este subconjunto especial de nuevas citas de la versión como revisiones con cambios en las conclusiones. Como todas las actualizaciones de las revisiones son muy importantes, aunque no cumplan con los criterios para una nueva cita de la versión, todas las revisiones actualizadas deben ser destacadas como revisiones actualizadas en la *CDSR* (por ejemplo, usando una etiqueta).

Los protocolos que tuvieron cambios importantes (por ejemplo, en los autores o en los criterios de selección) requieren una nueva cita de la versión. Los protocolos no aparecen en las bases de datos de referencias bibliográficas tales como MEDLINE e ISI, así que esto sólo afecta la cita que se hace en la CDSR. Los protocolos que cambian de tal manera que deben ser leídos de nuevo por los usuarios interesados requieren ser destacados en la CDSR (por ejemplo, usando una etiqueta). A estos los denominamos protocolos con un cambio importante.

Recuadro 3.2.a: Eventos que dan lugar a una Cita de la versión de un protocolo o revisión Cochrane

1. Un protocolo se publica por primera vez;

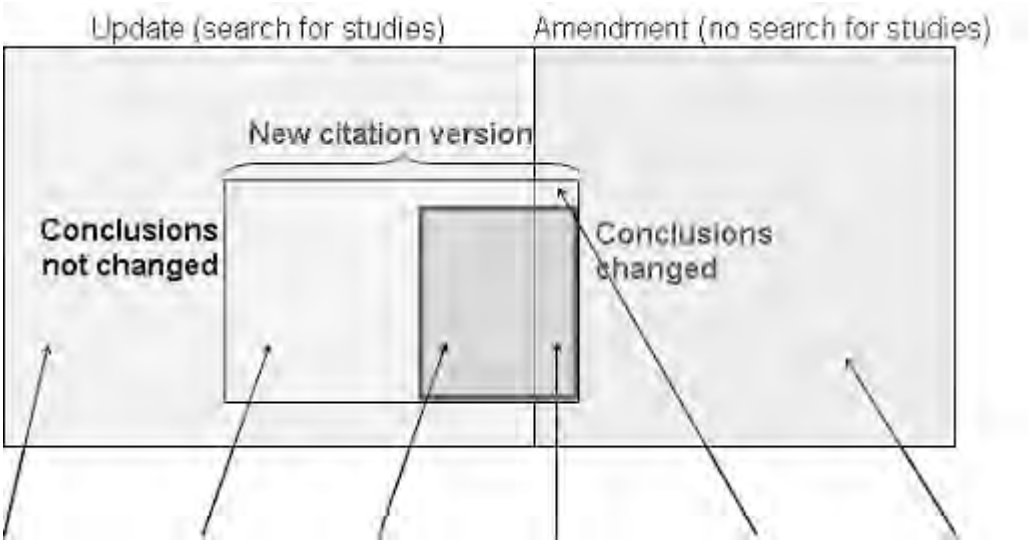
2. Un protocolo se publica de nuevo, después de considerarlo como una nueva cita de la versión;

3. Una revisión se publica por primera vez (es decir, cuando un protocolo se convierte en una revisión);

4. Una revisión se publica de nuevo (enmendada o actualizada) después de considerarla como una nueva cita de la versión;

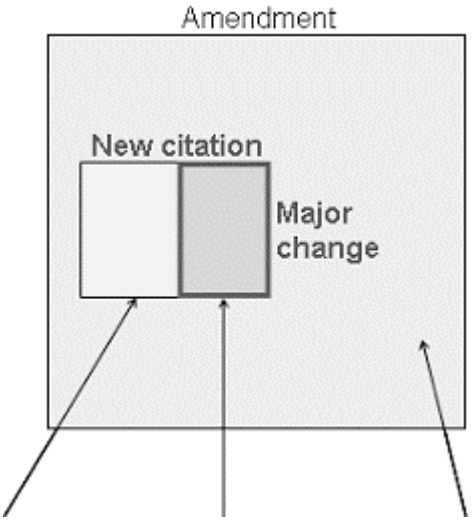
5. Una revisión es publicada de nuevo después de haber sido retirada; o una revisión es creada después de dividir un protocolo o una revisión ya existentes; o una revisión es creada después de fusionar protocolos o revisiones ya existentes.

Figura 3.2.a: Resumen de cambios en las revisiones Cochrane



Actualización, No hay una nueva cita *Por ejemplo, no hay cambios en las conclusiones o en los autores*	Actualización, Requiere una nueva cita, Sin cambios en las conclusiones *Por ejemplo, incluye cambio en los autores*	Actualización, Requiere una nueva cita, Cambios en las conclusiones *Por ejemplo, ahora hay suficiente evidencia de un efecto*	Enmienda, Requiere una nueva cita, Cambios en las conclusiones *Por ejemplo, corregir un error importante en las conclusiones (Erratas)*	Enmienda, Requiere una nueva cita, Sin cambios en las conclusiones *Por ejemplo, corregir un error importante de cita (Erratas)*	Enmienda, No requiere una nueva cita *Por ejemplo, corregir un pequeño error o cambio en los métodos*

Figura 3.2.b: Resumen de cambios en los protocolos Cochrane



Enmienda, Nueva cita, No hay cambios importantes en los planes *Por ejemplo, cambio de autores*	Enmienda, Nueva cita, Cambios importantes en los planes *Por ejemplo, cambio en el enfoque*	Enmienda, No requiere una nueva cita *Por ejemplo, corregir un pequeño error*

3.2.4 Aplicación de términos a los protocolos Cochrane

3.2.4.1 Enmiendas a los protocolos

Cualquier modificación o edición de un protocolo publicado le da al protocolo la condición de enmendado. No es posible 'actualizar' un protocolo. Los protocolos enmendados pueden volver a publicarse en la *CDSR*. Un protocolo puede ser enmendado en cualquier momento. Una enmienda puede significar mucho o poco trabajo, y producir cambios grandes o pequeños en el documento.

3.2.4.2 Nueva cita de la versión del protocolo

Un protocolo enmendado puede, a criterio del CRG, publicarse con una nueva cita de la versión, utilizando los criterios que se reflejan en el Recuadro 3.2.b. Esto cambia la forma de citación del documento dentro de la *CDSR*, aunque las citas de los protocolos no se incluyen en MEDLINE o en ISI.

Las nuevas citas de las versiones de los protocolos se clasifican, además, según si han sufrido un cambio importante o no. Un protocolo con un cambio importante será destacado en la *CDSR.*

Recuadro 3.2.b: Criterios para crear una nueva cita de la versión de un protocolo Cochrane

Criterios para crear una nueva cita de la versión de un protocolo: Con cambios importantes

Un protocolo debe ser clasificado como una nueva cita de la versión con cambios importantes si ha habido un cambio importante en los objetivos o en el alcance de la revisión propuesta, generalmente por medio de un cambio en los criterios de selección de los estudios para la revisión. Estos protocolos serán resaltados en la próxima publicación de la *CDSR*.

Criterios para crear una nueva cita de la versión de un protocolo: Sin cambios importantes

Un protocolo debe clasificarse como una nueva cita de la versión sin cambios importantes si ha tenido un cambio importante en el equipo de autores de la revisión. Estos protocolos no serán resaltados en la *CDSR.*

3.2.4.3 Ejemplos de cambios en los protocolos que no implican una nueva cita de la versión

Las siguientes enmiendas normalmente no deben conducir a que un protocolo sea clasificado como nueva cita de la versión, a menos que cumpla uno o ambos de los dos criterios enunciados en el Recuadro 3.2.b. Tales cambios se convertirán en una enmienda al protocolo publicado, pero las citaciones existentes se mantendrán.

- Cambios en el texto del protocolo (por ejemplo, la sección Antecedentes);
- Cambios en la metodología planificada;
- Cambios en el orden de los autores actuales (diferentes de un cambio en el primer autor), o supresión de autores;
- Correcciones.

3.2.5 Aplicación de términos a las revisiones Cochrane

3.2.5.1 Actualizaciones de las revisiones

Una actualización de una revisión Cochrane se define como cualquier modificación al documento publicado que incluye los hallazgos de una búsqueda de estudios adicionales (incluido el que no haya estudios nuevos), más reciente que los incluidos en la revisión publicada previamente. Se dice que la revisión ha sido actualizada. Cualquier nuevo estudio identificado debe incorporarse a la revisión actualizada (y no dejarlo dentro de los 'Estudios pendientes de clasificación', a menos que se hayan hecho todos los esfuerzos razonables para clasificarlo como estudios Incluidos, Excluidos o En marcha). También se considera una revisión como actualizada si una nueva búsqueda amplia no identificó ningún estudio adicional.

Esta definición se apoya en la definición de actualización de una revisión sistemática como "una acción puntual realizada con la intención de buscar e identificar nueva evidencia para incorporar a una revisión sistemática previamente completada" (Moher 2006). Una actualización de una revisión Cochrane puede significar mucho o poco trabajo, dependiendo de los resultados de la búsqueda, y debe hacerse al menos cada dos años (ver Sección 3.1.2).

3.2.5.2 Enmiendas a las revisiones

Una enmienda a una revisión Cochrane es cualquier modificación o edición que no incluya una actualización. Se dice entonces que la revisión ha sido enmendada. Ejemplos de enmiendas incluyen todos o algunos de los siguientes casos en ausencia de una nueva búsqueda de estudios: (1) un cambio en la metodología; (2) la corrección de un error de ortografía; (3) volver a escribir la sección de Antecedentes; (4) la inclusión completa de un estudio que figuraba como "pendiente de clasificación"; o (5) el cambio de conclusiones al descubrir un error importante de codificación de los datos. Una revisión Cochrane puede ser objeto de una enmienda en cualquier momento. Una enmienda puede suponer mucho o poco trabajo, y puede significar cambios grandes o pequeños para la revisión.

3.2.5.3 Nuevas citas de la versión de las revisiones

Una revisión Cochrane puede volver a publicarse con una nueva cita de la versión vigente. Sólo una actualización o una enmienda pueden tener esta posibilidad. Los autores y los CRG conjuntamente deciden si una revisión debe ser clasificada como una nueva cita de la versión. Hay seis criterios específicos para clasificar una revisión como nueva cita de la versión, que están descritos en el Recuadro 3.2.c. Salvo tres excepciones específicas (correcciones imprescindibles de las conclusiones, incorporación urgente de nueva información, y cambios imprescindibles en la cita de la revisión), sólo las versiones actualizadas son aptas para ser consideradas como una nueva cita de la versión.

Las nuevas citas de la versión vigente están, además, clasificadas como con 'Cambios en las conclusiones' o 'Sin cambios en las conclusiones'. Las revisiones con 'Cambios en las conclusiones' deben ser destacadas en la *CDSR*.

Las revisiones pueden ser actualizadas o enmendadas entre la publicación de nuevas citas de la versión, y estas revisiones actualizadas o enmendadas serán publicadas en la *CDSR* sin provocar una nueva cita. Por eso es tan importante que el grado en que una revisión haya sido actualizada quede bien reflejado en el apartado 'Fecha en que la revisión fue evaluada como actualizada' dentro de la revisión (ver Sección 3.3.2).

Recuadro 3.2.c: Criterios para crear una nueva cita de la versión de una revisión Cochrane

Criterios para crear una nueva cita de la versión de una revisión: Con cambios en las conclusiones

1. Cambio en las conclusiones de una actualización

Una revisión debe ser clasificada como nueva cita de la versión con cambios en las conclusiones si las Conclusiones de los Autores cambian durante una actualización hasta el punto de tener que recomendar a los usuarios que vuelvan a leer la revisión.

Estas conclusiones pueden cambiar como resultado de incorporar (o retirar) estudios, cambios en la metodología o cambios importantes del enfoque de la revisión (por ejemplo, nuevos desenlaces, comparaciones, tipos de participantes o cambios en la intervención o su modo de aplicación). Los cambios en las conclusiones tendrán casi invariablemente consecuencias en las implicaciones para la práctica en lo que se refiere a los efectos de la(s) intervención(es) estudiada(s). Sin embargo, algunas veces los cambios importantes tendrán implicaciones para la investigación (por ejemplo, si los nuevos datos incluidos resuelven incertidumbres que en la versión previa de la revisión se habían considerado como necesitadas de mayor investigación). Todos los cambios importantes en las conclusiones en

una revisión de con 'Cambios en las conclusiones' deben justificarse (y ser descritos) en el resumen de la revisión.

2. Cambio en las conclusiones al corregir un error grave (Fe de erratas)

Una revisión debe clasificarse como una nueva cita de la versión con cambios en las conclusiones si las Conclusiones de los Autores cambian tras la corrección de un error grave, a tal punto que se les deba recomendar a los usuarios que vuelvan a leer la revisión. Tales cambios son de aquellos que deberían justificar la publicación de una fe de erratas en una revista impresa tradicional.

3. Cambio en las conclusiones con incorporación urgente de nueva información acerca de los efectos de una intervención.

Una revisión debe clasificarse como una nueva cita de la versión con cambios en las conclusiones si las Conclusiones de los Autores cambian con la incorporación urgente de nueva información sobre los efectos de una intervención, a tal punto que se les deba recomendar a los usuarios que vuelvan a leer la revisión.

Criterios para crear una nueva cita de la versión de una revisión: Sin cambios en las conclusiones

4. Autoría nueva

Una revisión actualizada debe considerarse como una nueva cita de la versión sin cambios en las conclusiones, por decisión conjunta del CRG y de los autores, si se ha agregado una cantidad sustancial de información, o si ha habido cambios importantes en la metodología, o si la revisión ha tenido que ser realizada de nuevo en gran parte o ha sido escrita de nuevo (sin afectar las conclusiones) Y si ha habido un cambio importante en la lista de los autores (que incluya un cambio en el primer autor, lo que habitualmente no implica el reordenamiento o la supresión de otros autores), siempre que los autores cumplan los requisitos de autoría definidos en el Capítulo 4, Sección 4.2.2.

El compromiso que contraen los autores de las revisiones Cochrane de mantener su revisión puede requerir mucho trabajo para actualizar una revisión, y esto no implica que deban cambiar las conclusiones. Aunque el mismo equipo de autores de la revisión realice esfuerzos importantes, eso no debe dar lugar a una nueva cita de la versión si las conclusiones no cambian (ya que el equipo de autores de la revisión ya tiene la cita). Sin embargo, cuando el equipo de revisión cambia, mediante la adición o el reemplazo de autores, la revisión puede ser declarada como una nueva cita de la versión para dar el crédito apropiado a los nuevos autores.

5. Acumulación de cambios

Una revisión actualizada puede ser considerada una nueva cita de la versión sin cambios en las conclusiones, según criterio compartido del CRG y de los autores, si la cita de la versión tiene más de cinco años y la revisión ahora parece sustancialmente diferente de la versión citada, independientemente de cualquier cambio en las conclusiones o los autores. Una revisión puede parecer diferente, por ejemplo, debido a que fue escrita de nuevo, a la adición de numerosos estudios o debido a modificaciones sustanciales de la metodología que se hayan acumulado con el tiempo.

Debe tenerse en cuenta que cada revisión debe incluir una fecha, que corresponde a la fecha en que fue evaluada por última vez como actualizada. Por lo tanto, este criterio de considerar una revisión como una nueva cita de la versión debe utilizarse sólo para generar una nueva cita de la revisión en las bases de datos de referencias bibliográficas, tales como MEDLINE e ISI, y no para determinar la fecha en la cual ocurrieron los eventos o los cambios.

6. Corrección de errores importantes en la cita (Fe de erratas)

Una revisión puede clasificarse como una nueva cita de la versión sin cambios en las conclusiones si necesita cambiarse un error importante en las citas. Tales cambios (por ejemplo, la ortografía del nombre de un autor) son los detalles que justificarían la publicación de una fe de erratas en una revista impresa tradicional. Una actualización no es necesaria para una fe de erratas. Los errores importantes que afectan las conclusiones están definidos anteriormente en el criterio 2.

3.2.5.4 Ejemplos de cambios en las revisiones que no implican una nueva cita de la versión

Los siguientes cambios no deben en principio llevar a que una revisión sea clasificada como una nueva cita de la versión, a menos que la revisión también cumpla uno o más de los seis requisitos del Recuadro 3.2.c. Esos cambios pueden dar lugar a una actualización o a una enmienda de la revisión, pero la cita de la revisión existente se mantendrá.

- La adición de nuevos estudios en los que no se produzca un cambio sustancial en la revisión.
- Cambios en los resultados de los análisis (por ejemplo, en cálculos del efecto o intervalos de confianza), sin cambio en las conclusiones;
- Cambios en el texto de la revisión (por ejemplo, las secciones de Antecedentes o Discusión);
- Cambios en la metodología;
- Cambios en el orden de los autores actuales (diferente de cambiar el primer autor), o supresión de autores;
- Correcciones.

3.3 Fechas importantes asociadas con las revisiones Cochrane

3.3.1 Introducción

Hay varias fechas asociadas con una revisión Cochrane. Algunas son producidas automáticamente por RevMan, y otras deben ser introducidas por el autor de la revisión. Estas fechas son importantes tanto para informar a los lectores de la revisión como para facilitar el manejo de la publicación de la revisión. Es importante que los autores apliquen estas definiciones cuando introduzcan fechas en campos relevantes durante la actualización o la enmienda de una revisión.

3.3.2 Fecha en que la revisión fue evaluada como actualizada

Debe ser introducida por uno de los autores de la revisión (sólo en caso de las revisiones, no en los protocolos). En la publicación, esta fecha aparece en un sitio destacado de la revisión para informar a los lectores cuán recientemente ha sido evaluada la revisión como actualizada. Los criterios para evaluar una revisión como actualizada aparecen en la Tabla 3.3.a.

Una revisión debe considerarse actualizada aún cuando haya recibido sólo mínimas correcciones en muchos años, por ejemplo, si una búsqueda reciente de estudios encuentra que no hay nueva evidencia desde que fue publicada la revisión. Todas las revisiones admitidas para su publicación deben incluir la última fecha en que la revisión fue evaluada como actualizada. La fecha debe ser introducida por los autores, y a menudo coincide con la fecha en que los autores envían la revisión para considerar su publicación en la CDSR. Es conveniente corregir la fecha en que la revisión es aceptada para publicarla.

Recuadro 3.3.a: Guía para declarar actualizada una revisión

La fecha en que una revisión es evaluada como actualizada debe elegirse de tal manera que la revisión (nueva, actualizada o enmendada) cumpla los siguientes criterios clave:

1. La evidencia está actualizada sobre los efectos de la(s) intervención(es)

La lista de estudios incluidos debe reunir toda la evidencia disponible, la cual debe resultar de la búsqueda realizada más recientemente, preferentemente dentro de los seis meses en los que la revisión es evaluada como actualizada;
Además, es muy deseable, pero no obligatorio, que:

2. Los métodos de la revisión estén actualizados

Deben estar incorporados todos los métodos obligatorios para las revisiones Cochrane (como se describe en la versión vigente del *Manual Cochrane para Revisiones Sistemáticas de Intervenciones*);

3. Las afirmaciones basadas en hechos son correctas

Las afirmaciones basadas en hechos, por ejemplo, en los Antecedentes y la Discusión, no deben estar excesivamente desactualizadas.

3.3.3 Fecha de la búsqueda

Esta fecha debe ser introducida por los autores de la revisión (para revisiones solamente, no para protocolos). 'Búsqueda' se refiere aquí a las búsquedas de todas las bases de datos consultadas para la revisión. Si se consultaron varias bases de datos en diferentes fechas, en la revisión debe anotarse la fecha más reciente de búsqueda de cada una de las bases de datos, y la primera de las fechas debe anotarse en esta casilla. Por ejemplo, si las búsquedas más recientes de las siguientes bases de datos fueron en las siguientes fechas (MEDLINE 5 de Junio de 2007, EMBASE 12 de Junio de 2007, Registro Especializado de los CRG 26 de Junio de 2007 y CENTRAL el 28 de Junio de 2007), la 'Fecha de búsqueda' sería el 5 de Junio de 2007.

3.3.4 Fecha esperada de la próxima etapa

Introducida por los autores de la revisión como:

- Para los protocolos: la fecha en la que se espera la revisión completa;
- Para las revisiones: la fecha en la que se espera la próxima actualización.

3.3.5 Fecha de la última corrección

Esta fecha se registra automáticamente en RevMan, a partir de cualquier modificación de la revisión, y no será publicada. Se utilizará para determinar la fecha en la que la actual revisión publicada apareció primero, tal y como es.

3.3.6 Fecha en que se declaró que la revisión no necesita más actualizaciones

Esta fecha se aplica a muy pocas revisiones y debe ser empleada con precaución, después de ser consultada con el Grupo Cochrane de Revisión (CRG). Una revisión que no se actualizará más es la que tiene altas posibilidades de mantener su relevancia actual en un futuro inmediato (medido en años más que en meses). Tales revisiones son más la excepción que la regla, y la decisión de no realizar más actualizaciones a una revisión debe consultarse con el CRG, y revisarse periódicamente. Las situaciones en que se decide que una revisión no requiere más actualizaciones incluyen:

- La intervención ha quedado relegada (teniendo en cuenta que las revisiones Cochrane deben ser relevantes desde una perspectiva internacional);
- La conclusión es tan firme que la incorporación de nueva información no la cambiará, y no hay efectos adversos previsibles de la intervención.

La revisión permanece como 'no más actualizaciones' mientras la entrada más reciente de 'Novedades' la declare como tal. Si se incorpora posteriormente una declaración en 'Novedades', se considerará entonces que la revisión debe ser actualizada, como cualquier otra de las revisiones Cochrane.

3.4 Consideraciones cuando se actualiza una revisión Cochrane

3.4.1 Por dónde comenzar

Se han realizado pocos estudios metodológicos para tomar decisiones acerca de cómo y cuándo deberían actualizarse las revisiones sistemáticas (Moher 2008). Sin embargo, este es un tema de creciente interés y las directrices que contiene este capítulo serán actualizadas regularmente de acuerdo con los nuevos conocimientos en investigación metodológica.

La actualización de una revisión Cochrane generalmente debe hacerse cada dos años y debe incluir una búsqueda de nuevos estudios. Si se identifican nuevos estudios, deben evaluarse para su inclusión, y si se consideran elegibles, deben incorporarse a la revisión. Mientras se prepara la actualización de una revisión, deben tenerse en cuenta otros asuntos adicionales, por ejemplo:

1. Cualquier necesidad de cambiar la pregunta de investigación y los criterios de selección de los estudios para la revisión: por ejemplo, la incorporación de una nueva variable de resultado o comparación, la incorporación de un nuevo análisis específico de subgrupos, la incorporación de métodos mejorados para categorizar la condición;

2. Cambios en Metodología: por ejemplo, la inclusión de la evaluación del 'Riesgo de sesgo' de los estudios actualmente incluidos (Capítulo 8) o la adición de una tabla de 'Resumen de los hallazgos' (Capítulo 11).

3.4.2 Actualización de una revisión con una pregunta de revisión sin cambios

3.4.2.1 Ejecución de una nueva búsqueda

Cuando no ha habido cambios en la pregunta de revisión o en los criterios de selección de los estudios, la búsqueda de nuevos estudios es el primer paso y definitivo del proceso de actualización. Para los CRG con suficientes recursos, la identificación periódica de estudios potencialmente relevantes y el envío de referencias bibliográficas a los autores de la revisión es una función actual del equipo editorial (usualmente es el papel del Coordinador de la Búsqueda de Ensayos). En otras ocasiones, los autores de la revisión deberán hacer ellos mismos la búsqueda. Por lo menos, las estrategias para identificar nuevos estudios para actualizar una revisión deben incluir una nueva ejecución de la estrategia de búsqueda, desde la 'Fecha de búsqueda' de la última actualización (ver Capítulo 6, Sección 6.4.12).

Siempre que ocurren grandes avances en los métodos de búsqueda, o cuando los autores de la revisión crean que la estrategia de búsqueda de la revisión original podría mejorarse, la nueva búsqueda necesitará hacerse para el período que empieza en la fecha de la última búsqueda, usando los términos de búsqueda adicionales o modificados y aplicándolos al período de búsqueda cubierto en la revisión original.

3.4.2.2 Actualización de las revisiones cuando no se encuentran nuevos estudios

Cuando no se encuentran nuevos estudios que cumplan los criterios de selección, la actualización de la revisión simplemente requiere que este hallazgo sea registrado en las secciones correspondientes de la revisión. La revisión del texto de la revisión puede ser necesaria en las siguientes secciones:

1. Los métodos de búsqueda (para estar seguros, se debe registrar la 'Fecha de búsqueda' correcta);

2. La descripción de los estudios en la sección de Resultados (para revisar el número de los estudios identificados, seleccionados y excluidos, si es el caso);

3. Resultados (para estar seguros que los datos son correctos);

4. Conclusiones de los Autores de la Revisión (especialmente si todavía existe la necesidad de realizar más estudios);

5. Resumen y resumen en lenguaje sencillo

Además de revisar el texto de la revisión, los autores deberán asegurar que los campos relacionados con la fecha son correctos y reflejan el estado de actualización de la revisión (ver Sección 3.3), y que la tabla 'Qué hay de nuevo' se ha completado (ver Sección 3.5).

Con el fin de advertir a los lectores de la revisión que están leyendo una versión actualizada, se debe añadir una frase a la sección de Antecedentes del Resumen, manifestando que es una actualización de una versión Cochrane (con la cita de la versión original) y señalando el año en que se publicó inicialmente la revisión junto con las fechas de cualquiera de las actualizaciones previas. En la sección Antecedentes de la revisión misma, esta frase puede ampliarse para incluir la discusión de los hallazgos de la revisión original.

Finalmente, es importante comprobar que no hay nada más desactualizado en la revisión (por ejemplo, referencias a otras revisiones Cochrane que pueden haber sido actualizadas, información acerca de la prevalencia o incidencia de la condición de interés, declaraciones como 'recientemente, en 1998, se demostró que….', 'en el próximo año 2002 habrá…'). Si hay cambios o adiciones en las secciones Agradecimientos y Potenciales Conflictos de Interés de la revisión, estos deben revisarse.

3.4.2.3 Actualización de revisiones cuando se encuentran nuevos estudios

Si se encuentran nuevos estudios potencialmente relevantes, deben ser evaluados para incluirlos en la revisión utilizando el mismo proceso (y formulario de selección de estudios) de la revisión original (para información sobre la selección de estudios, ver Capítulo 5).

Si se decide incluir nuevos estudios en la revisión actualizada, se deberán introducir las referencias bibliográficas en RevMan, extraer los datos (ver Capítulo 7) y evaluar el riesgo de sesgo (ver Capítulo 8). Los datos recogidos de los nuevos estudios identificados e incluidos deben introducirse en RevMan y, si es razonable, se debe realizar un (nuevo) metanálisis (Capítulo 9). En lo posible, los métodos utilizados en la actualización de la revisión deberían reproducir los métodos de la revisión original, a menos que se hayan alterado explícitamente (por ejemplo, mediante nuevos métodos desarrollados para las revisiones sistemáticas, tales como el uso de tablas de 'Riesgo de sesgo', o la inclusión de tablas 'Resumen de hallazgos'). En los casos en que los métodos difieran de los utilizados en la revisión original, estas diferencias y su justificación deberán documentarse en la sección de la revisión 'Diferencias entre la revisión y el protocolo'.

La cantidad de revisión requerida en el texto de una revisión actualizada que incluye nuevos estudios dependerá de la influencia de los nuevos datos en los resultados de la revisión. Los ejemplos van desde la adición de estudios pequeños que no traen cambios en los resultados o las conclusiones de la revisión (y por lo tanto requieren muy poca revisión del texto más allá de lo descrito en la Sección 3.4.2.2), aumentando así la certeza de los resultados y las conclusiones previas (que requieren alguna modificación del texto) y, en algunos casos, un cambio en la conclusión de una revisión (con la necesidad posterior de volver a escribir más ampliamente los Resultados, la Discusión, la Conclusión, la tabla de 'Resumen de hallazgos', el Resumen y el resumen en lenguaje sencillo). Además, debería incluirse una nota en las secciones Resumen y Antecedentes de la revisión, donde se advierta a los lectores del hecho de estar leyendo una versión actualizada de una revisión previa (3.4.2.2).

Los autores deberán asegurar que las fechas de los campos relevantes son correctas y reflejan el estado actualizado de la revisión (ver Sección 3.3.2) y que la Tabla 'Novedades' está completada (ver Sección 3.5). Finalmente, los autores deberían revisar que nada más en la revisión esté desactualizado (por ejemplo, las referencias a otras revisiones Cochrane que pueden haber sido actualizadas, la información sobre la prevalencia o la incidencia de la condición de interés, declaraciones como 'recientemente, en 1998, se demostró que…', 'el próximo año 2002, habrá…'). Si hay cambios o adiciones en las secciones Agradecimientos y Potenciales Conflictos de Intereses de la revisión, deberán ser revisados.

3.4.3 Cambios en la pregunta de la revisión y los criterios de selección

Hay ocasiones en que, además de volver a realizar la búsqueda, una actualización de una revisión incluye también un cambio en la pregunta de la revisión, en los criterios de selección de los estudios, o en ambos. Por ejemplo, la cambiante tecnología puede llevar a la inclusión de una nueva comparación; o una categoría de pacientes (por ejemplo, niños además de adultos), o un resultado importante (por ejemplo, efectos adversos), que no ha sido acordado adecuadamente en la revisión original. Si este es el caso, los cambios y las adiciones propuestas al protocolo original deberían ser documentadas y justificadas en la sección 'Diferencias entre el protocolo y la revisión', explicado en el texto de la revisión (Secciones de Antecedentes, Objetivos y Métodos) y destacados en la tabla 'Qué hay de nuevo'.

Además, los métodos de búsqueda pueden necesitar ser cambiados y realizados de nuevo para que cubran no sólo el período desde la 'Fecha de búsqueda' de la versión previa de la revisión, sino también el período cubierto por la revisión original con la adición de nuevos términos de búsqueda relevantes para cualquier criterio de selección adicional. En algunos casos es suficiente con volver a los resultados de búsqueda originales y aplicar los criterios de selección para la inclusión de estudios ya actualizados.

Si se ha agregado a la revisión una nueva comparación o un nuevo resultado, será necesario volver a los estudios originales incluidos y verificar que no incluyen alguna información relevante para este nuevo resultado o comparación. Los formularios originales de extracción de datos pueden necesitar ser cambiados o ampliados y probados de nuevo, y deben añadirse al análisis las nuevas comparaciones o los nuevos resultados.

Finalmente, la adición de nuevas comparaciones, poblaciones o resultados significará la necesidad de modificar el texto de la revisión (Antecedentes, Métodos), y si se identifican y se incluyen estudios adicionales, debe hacerse lo mismo en los Resultados, Conclusiones, Resumen en lenguaje sencillo y la tabla de 'Resumen de hallazgos'.

3.4.4 Partición de Revisiones

Muchas veces una revisión puede volverse muy extensa y puede ser deseable partir la revisión en dos o más revisiones nuevas. Partir las revisiones en temas de revisión definidos de forma más restringida, posiblemente con menos estudios, puede facilitar la actualización y permite compartir el esfuerzo de actualización entre varios equipos de revisión.

La partición de una revisión implica crear al menos una nueva cita de la versión de la revisión, y puede perderse el enlace formal con las versiones previas. Partir una revisión muchas veces incluye retirar la revisión original. La decisión de partir una revisión no debe hacerse a la ligera y siempre deber ser consultado con el grupo editorial del GRC.

Las revisiones globales (*overviews*) Cochrane (ver Capítulo 22) pueden facilitar la partición de las revisiones con la posibilidad de tener varias revisiones definidas más específicamente (por ejemplo, intervenciones únicas para una condición particular), combinándose en una revisión global (*Overview*) de todas las intervenciones para esa misma condición sanitaria particular.

3.4.5 Actualización de la metodología de una revisión

Además de la búsqueda de nuevos estudios y de revisar la pregunta de la revisión o los criterios de selección de estudios, el mantenimiento de una revisión puede incluir la modificación de la metodología de la revisión (Shea 2006). Los avances metodológicos en la realización de una revisión sistemática desde la fecha de publicación de la revisión original pueden hacer necesario revisar o ampliar los métodos de la revisión durante una actualización. Los autores de la revisión pueden decidir incluir una nueva estrategia de análisis en su revisión actualizada (por ejemplo, utilizando métodos estadísticos no disponibles previamente en RevMan). La introducción de las tablas 'Riesgo de sesgo' (Capítulo 8) y 'Resumen de hallazgos' (Capítulo 11) con RevMan, mientras no sea obligatorio, dan la oportunidad de que las revisiones se actualicen para incluir estos nuevos métodos. Cuando se introduce una tabla de 'Riesgo de sesgo' a una revisión, los autores deben decidir si volver a revisar la evaluación crítica de los estudios incluidos en versiones previas de la revisión, actualizando todas las evaluaciones del riesgo de sesgo, o aplicar estos métodos nuevos sólo a los estudios agregados en la actualización. En la versión publicada de la revisión debería elaborarse una tabla de 'Riesgo de sesgo' que incluya sólo aquellos estudios donde se hayan introducido datos (es decir, sin filas en blanco).

Como parte de la actualización de una revisión, los autores pueden querer incluir una tabla de 'Resumen de hallazgos' (Capítulo 11). Los desenlaces seleccionados para ser presentados en la tabla 'Resumen de hallazgos' deben ser los que tengan importancia para aquellos que toman decisiones sanitarias (generalmente los desenlaces primarios de la revisión), y deben seleccionarse antes de comenzar la actualización para reducir el riesgo de informar selectivamente de resultados significativos en vez de informar sobre aquellos de importancia.

Los cambios en la metodología pueden implicar cambios en el protocolo original de la revisión. Estos cambios, y su justificación, deben registrase explícitamente en la sección 'Diferencias entre el protocolo y la revisión' y en la Tabla 'Novedades'.

3.4.6 Otros cambios en la revisión

Si hay un cambio del autor principal, se han agregado nuevos autores al equipo, o un nuevo grupo ha hecho la revisión, la lista de autores necesita ser cambiada. La decisión de quién es nombrado en la lista de autores de una revisión actualizada, y en qué orden, debe relacionarse con las contribuciones históricas a la revisión actualizada junto con el visto bueno del documento final actualizado. Si un autor no está dispuesto a aprobar una revisión actualizada, ese autor no debe aparecer mencionado en la lista de autores, pero debe ser mencionado en los Agradecimientos. Las contribuciones de todos los autores, tanto de la versión inicial como de la actualizada, deben ser inscritas en la sección 'Contribuciones de los autores'.

Cambiar los autores de una revisión puede implicar tener que concederle a la revisión una nueva cita de la versión (ver Sección 3.2.5.3).

3.4.7 Proceso editorial

Después de completar el proceso de actualización, la revisión debe ser enviada al equipo editorial para continuar el proceso. Hay variación en las políticas de los CRG sobre si las revisiones actualizadas pueden continuar el proceso completo de revisión editorial y cuándo pueden hacerlo. Si una actualización no incluye un amplio análisis o cambio del resultado, no necesita ser remitida; sin embargo, si hay análisis nuevos, inclusión de nuevos métodos o cambios en las conclusiones, tanto el proceso de prepublicación como el de la revisión original debe repetirse.

En raras ocasiones una revisión debe retirarse de la CDSR. Esto puede ser temporal (por ejemplo, porque la revisión está muy desactualizada, o contiene un grave error), o permanente (por ejemplo, porque la revisión ha sido partida en una serie de revisiones más pequeñas). La retirada de la revisión debe anotarse en la sección 'Notas publicadas' de la revisión. La revisión que contiene esta nota de retirada debe ser aceptada para publicación en cada número de la CDSR. Si la retirada es temporal, la revisión debe ser repuesta cuando los autores de la revisión y el CRG decidan que el contenido es satisfactorio. Si una revisión es retirada porque su contenido ha sido incluido en otra revisión, debe darse la noticia en la sección 'Notas publicadas', para explicar que su retirada ha sido por esta razón.

3.5 Tablas 'Novedades' e Historial

3.5.1 Los eventos en 'Novedades'

Todas las revisiones y los protocolos actualizados y enmendados deben generar una tabla de 'Qué hay de nuevo', de tal manera que los lectores puedan, de forma rápida y clara, identificar lo que ha cambiado. Los eventos introducidos a la tabla 'Qué hay de nuevo' determinan qué categoría tiene el protocolo o la revisión en la CDSR, incluido el uso de etiquetas u otros medios para destacarlos y asignarles una nueva cita de la versión.

3.5.2 Completando la tabla 'Qué hay de nuevo'

Cada casilla de una tabla 'Novedades' o Historial incluye:

- La fecha en la que se realizó o registró el evento;
- El tipo de evento;
- Una descripción breve de los cambios que se hicieron

La tabla 3.5.a y la tabla 3.5.b enumeran los eventos de la tabla 'Qué hay de nuevo' disponibles en los protocolos y en las revisiones, respectivamente. Los autores deben remitirse a la sección de referencia con tal de seleccionar el evento apropiado para incluirlo en la tabla 'Qué hay de nuevo'.

Tabla 3.5.a: Eventos posibles para la tabla 'Novedades' de los protocolos

Tipo de episodio	Definición o discusión	Implicación para el protocolo publicado
Enmendado.	Ver 3.2.2 y 3.2.4.1.	Ninguno.
Comentario incorporado.	Ver 3.6.	Protocolo destacado como 'comentario incorporado'.
Nueva cita: sin cambios importantes.	Nueva cita: sin cambios importantes.	Nueva cita.
Nueva cita: cambios importantes.	Ver 3.2.4.2.	Nueva cita. Protocolo destacado como 'cambios importantes'

Tabla 3.5.b: Eventos posibles para la tabla 'Novedades' de las revisiones

Tipo de episodio	Definición o discusión	Implicación para la revisión publicada
Enmendada.	Ver 3.2.2 y 3.2.5.2.	Ninguno.
Actualizada.	Ver 3.2.2 y 3.2.5.1.	Revisión señalada como actualizada.
Comentario incorporado.	Ver 3.6.	Revisión señalada como 'comentario incorporado'
Nueva cita: sin cambios en las conclusiones.	Ver 3.2.3 y 3.2.5.3.	Nueva cita (por ejemplo, registro en MEDLINE); Contador del

		factor de impacto de los cambios señalados.
Nueva cita: con cambios en las conclusiones.	Ver 3.2.3 y 3.2.5.3.	Revisión clasificada como con 'Cambios en las conclusiones'. Nueva cita (por ejemplo, registro en MEDLINE); Contador del factor de impacto de los cambios señalados.
No más actualizaciones.	Ver 3.3.6.	Ninguno.

Aunque técnicamente es posible introducir varios eventos en la tabla 'Qué hay de nuevo', los autores deben asegurarse que la tabla sólo incluya información de los cambios producidos desde la última cita de la versión. Principalmente, la tabla no debe tener más de una entrada de nueva cita de la versión o más de una entrada de actualización (los eventos previos deben trasladarse a la tabla Historial).

3.5.3 Tabla de Historial

Las entradas a la tabla 'Qué hay de nuevo' deben trasladarse a la tabla de Historial cuando ya no pertenezcan a la última cita de la versión del protocolo o de la revisión. Además, la tabla de Historial incluirá la siguiente información, que deberá ser completada automáticamente por el sistema de gestión de información de la Colaboración.

- Año y número de la primera publicación del protocolo;
- Año y número de la primera publicación de la revisión;
- Año y número de cada nueva cita de la versión.

3.6 Incorporar y orientar la retroalimentación o comentarios a una revisión Cochrane

Existe un mecanismo establecido por la Colaboración Cochrane para agilizar y manejar la retroalimentación o comentarios procedentes de los usuarios de las revisiones. La retroalimentación, antes denominada Comentarios y Críticas, está diseñada para "… corregir las revisiones a la luz de nueva evidencia…. para reflejar la aparición de nuevos datos, o cualquier otra retroalimentación válida, solicitada o no, de cualquier origen" (Chalmers 1994).

La retroalimentación a una revisión puede recibirse en cualquier momento después de la publicación, y se enviará al editor de Retroalimentación del CRG correspondiente. Este editor se asegurará de que la retroalimentación o comentarios recibidos y el lenguaje sean apropiados y lo pasará a los autores para su revisión y respuesta (generalmente al mes de haber sido enviada). Al responder a la retroalimentación los autores deben:

- Limitar la respuesta a los puntos mencionados en la retroalimentación;
- Responder a cada punto principal, estableciendo explícitamente cuándo el autor está de acuerdo o no con la retroalimentación y aportar pruebas de evidencia cuando sea necesario;
- Describir cualquiera de los cambios hechos a la revisión como respuesta a la retroalimentación;
- Responder en lenguaje claro y sencillo.

Actualizar una revisión da la oportunidad de incorporar la retroalimentación o comentarios recibidos en la revisión, abordando asuntos válidos y agregando cualquiera de los estudios adicionales identificados a través de los mecanismos de retroalimentación.

3.7 Información del capítulo

Autores: Julian PT Higgins, Sally Green y Rob JPM Scholten.

La versión en inglés de este capítulo debe ser citada como: Higgins JPT, Green S, Scholten RJPM. Chapter 3: Maintaining reviews: updates, amendments and feedback. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: El Grupo de Trabajo en Actualizaciones de la Colaboración Cochrane (integrado por Mike Clarke, Mark Davies, Davina Ghersi, Sally Green, Sonja Henderson, Harriet MacLehose, Jessie McGowan, David Moher, Rob Scholten (coordinador) y Phil Wiffen) aportaron comentarios a los borradores.

3.8 Referencias

Chalmers 1994
Chalmers I, Haynes B. Reporting, updating, and correcting systematic reviews of the effects of health care. *BMJ*. 1994;309:862-865.

Chapman 2002
Chapman A, Middleton P, Maddern G. Early updates of systematic reviews - a waster of resources? Pushing the Boundaries: Fourth Symposium on Systematic Reviews, Oxford, 2002.

Moher 2006
Moher D, Tsertsvadze A. Systematic reviews: when is an update an update? *The Lancet.* 2006;367:881-883.

Moher 2007
Moher D, Tsertsvadze A, Tricco AC, Eccles M, Grimshaw J, Sampson M, Barrowman N. A systematic review identified few methods and strategies describing when and how to update systematic reviews. *Journal of Clinical Epidemiology.* 2007;60:1095-1104.

Moher 2008
Moher D, Tsertsvadze A, Tricco AC, Eccles M, Grimshaw J, Sampson M, Barrowman N. When and how to update systematic reviews. *Cochrane Database of Systematic Reviews* 2008, Issue 1. Art No: MR000023.

Shea 2006
Shea B, Boers M, Grimshaw JM, Hamel C, Bouter LM. Does updating improve the methodological and reporting quality of systematic reviews? *BMC Medical Research Methodology*. 2006;6:27.

Shojania 2007a
Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Annals of Internal Medicine.* 2007;147:224-233.

Shojania 2007b
Shojania KG, Sampson M, Ansari MT, Ji J, Garritty C, Rader T, Moher D. *Updating Systematic Reviews. Technical Review No 16 (Prepared by the University of Ottawa Evidence-based Practice Center under Contract No 290-02-0017)*. Rockville (MD): Agency for Healthcare Research and Quality, 2007.

ÍNDICE
CAPÍTULO 4: GUÍA DEL CONTENIDO DE UN PROTOCOLO Y UNA REVISIÓN COCHRANE

PUNTOS CLAVE

4.1 INTRODUCCIÓN

4.2 TÍTULO E INFORMACIÓN DE LA REVISIÓN (O INFORMACIÓN DEL PROTOCOLO)
- 4.2.1 Título
- Tabla 4.2.a: Estructura para los títulos de las revisiones Cochrane
- 4.2.2 Los autores
- 4.2.3 Persona de contacto
- 4.2.4 Fechas
- 4.2.4.1 Evaluada como actualizada
- 4.2.4.2 Fecha de búsqueda
- 4.2.4.3 Etapa siguiente
- 4.2.4.4 Primera publicación del protocolo
- 4.2.4.5 Primera publicación de la revisión
- 4.2.4.6 El volumen de la última cita
- 4.2.5 Novedades e Historia

4.3 RESUMEN

4.4 RESUMEN EN TÉRMINOS SENCILLOS

4.5 TEXTO PRINCIPAL

4.6 TABLAS
- 4.6.1 Características de los estudios incluidos
- 4.6.2 Riesgo de sesgo
- 4.6.3 Características de los estudios excluidos
- 4.6.4 Características de los estudios pendientes de clasificación
- 4.6.5 Características de los estudios en marcha
- 4.6.6 Resumen de los hallazgos
- 4.6.7 Tablas adicionales

4.7 ESTUDIOS Y REFERENCIAS
- 4.7.1 Referencias a los estudios
- 4.7.2 Otras referencias

4.8 DATOS Y ANÁLISIS
- Figura 4.8.a: Ilustración de la jerarquía de la sección ‘Datos y análisis’.

4.9 FIGURAS
- Tabla 4.9.a: Tipos de figuras que pueden incluirse en una revisión Cochrane
- 4.9.1 Gráficos del RevMan
- 4.9.2 Otras figuras

4.10 FUENTES DE FINANCIACIÓN DE LA REVISIÓN

4.11 COMENTARIOS

4.12 ANEXOS

4.13 INFORMACIÓN DEL CAPÍTULO

4.14 REFERENCIAS

CAPÍTULO 4: GUÍA DEL CONTENIDO DE UN PROTOCOLO Y UNA REVISIÓN COCHRANE

Editores: Julian PT Higgins y Sally Green.

Puntos clave

- Las revisiones Cochrane tienen un formato muy estructurado. Conforme a ello, este formato es facilitado por el uso de RevMan. Este capítulo describe qué se espera que incluya un autor, y qué puede esperar encontrar un lector en cada componente de un protocolo o de una revisión Cochrane;

- El capítulo también sirve como una guía para gran parte del Manual, al contener enlaces con otros capítulos donde se puede hallar una discusión en profundidad sobre los aspectos metodológicos;

- Una sección de información sobre la Revisión (o una sección de 'Información sobre el Protocolo') incluye detalles sobre los autores y fechas importantes asociadas con el mantenimiento y la actualización de de la revisión;

- El texto principal debe ser sucinto y legible, de tal modo que alguien que no sea experto en el tema pueda comprenderlo. El texto de un protocolo termina después de la sección de Métodos;

- Una sección de 'Estudios y referencias' aporta un marco de clasificación tanto de los estudios incluidos, excluidos y en realización, como de aquellos que no tienen suficiente información disponible, así como de otras referencias;

- Las tablas de características de los estudios permiten la presentación sistemática de los descriptores clave de los estudios considerados para la revisión;

- Una sección de 'Datos y Análisis' tiene una estructura jerárquica, que así permite que los datos de los estudios incluidos sean colocados dentro de subgrupos particulares de estudios, los que a su vez están dentro de metanálisis de variables de resultado específicas, que a su vez están dentro de comparaciones de intervenciones específicas. Para cada metanálisis se pueden construir gráficos de bosque y gráficos de embudo dentro del RevMan;

- Se pueden añadir tablas, figuras y anexos para complementar las tablas por defecto.

4.1 Introducción

Todas las revisiones Cochrane de intervenciones tienen el mismo formato, y la preparación de una revisión en el formato requerido se consigue con el uso del software RevMan. En este capítulo discutimos el contenido completo de la revisión (o del protocolo) y definimos qué debe aparecer en cada sección. Se incluyen amplias referencias a otros capítulos del Manual para identificar consejos relevantes a cada sección. La guía para utilizar el software RevMan está disponible en el sistema de ayuda dentro del mismo programa.

4.2 Título e información de la revisión (o información del protocolo)

4.2.1 Título

El título define brevemente la intervención(es) revisada(s) y el problema al cual se dirige la intervención. La guía explícita para estructurar los títulos de las revisiones Cochrane se presenta en la Tabla 4.2.a.

Tabla 4.2.a: Estructura para los títulos de las revisiones Cochrane

Escenario	Estructura	Ejemplo
Estructura básica.	[Intervención] para [problema sanitario].	Antibióticos para la bronquitis aguda.
Comparar dos intervenciones activas.	[Intervención A] frente a [intervención B] para [problema sanitario].	Tratamiento inmediato frente a tratamiento diferido de la neoplasia intraepitelial cervical.
Tipo de persona estudiada o lugar de intervención explícitamente mencionados.	[Intervención] para [problema sanitario] en [grupo participante/lugar].	Óxido nítrico inhalado para el fallo respiratoria en los bebés prematuros.
No especificar un 'problema sanitario' en particular (por ejemplo, 'parto en casa versus parto en hospital'), o si la intervención pretende influir en una variedad de problemas (por ejemplo, 'Sufactante sintético profiláctico en bebés prematuros").	[Intervención] en quirófano para [grupo participante /lugar].	Ingesta de agua restringida frente a ingesta libre en bebés prematuros.
En ocasiones es necesario especificar que la intervención es para prevenir, para tratar, o para prevenir y tratar el(los) problema(s) sanitario(s): Si es necesario, la palabra 'para' va seguida de 'prevenir', 'tratar', o 'prevenir y tratar'. Esto es mejor que decir 'para la prevención de' etc.		Para prevenir que los niños se ahoguen se deben utilizar vallas en la piscina; Amodiaquina para el tratamiento de la malaria; Vitamina C para la prevención y el tratamiento del resfriado común.

4.2.2 Los autores

La autoría de todos los artículos científicos (incluidos los protocolos y las revisiones Cochrane) implica obligación, responsabilidad y crédito (Rennie 1997, Flanagin 1998, Rennie 1998). Al decidir quién debe aparecer en la lista de autores de una revisión Cochrane, es importante distinguir a quienes han hecho una contribución sustancial a la revisión (los cuales deben aparecer en esta lista) y aquellos que han ayudado de otras formas, los cuales deben ser mencionados en la sección de Agradecimientos. La autoría debe basarse en contribuciones sustanciales en la totalidad de la tres etapas siguientes, según 'Los requerimientos homogéneos para los manuscritos presentados a las revistas médicas' (Comité Internacional de Editores de Revistas Médicas 2006). Los autores deben firmar un formato de 'Licencia de publicación' que establece las siguientes tres contribuciones:

- Idea y diseño del estudio, o análisis e interpretación de los datos
- Redactar la revisión o comentar críticamente el contenido intelectual
- Aprobación final del documento que se va a publicar.

Las contribuciones específicas deben mencionarse en la sección 'Contribución de los autores' (ver más abajo). La lista de autores puede ser el nombre de una persona, varias personas, un grupo colaborador (por ejemplo, 'Grupo del Estudio del Cáncer Vesical Avanzado') o una combinación de uno o más autores y un grupo colaborador. Idealmente, el orden de los autores debe ser proporcional a sus contribuciones relacionadas con la revisión. La persona que más ha contribuido debe encabezar la lista.

4.2.3 Persona de contacto

Se deben proporcionar los detalles de contacto de la persona a quien se le debe dirigir la correspondencia sobre la revisión, la cual habrá aceptado tomar la responsabilidad de mantener y desarrollar la revisión. La mayoría de las veces esta persona debería (i) responsabilizarse del desarrollo y la organización del equipo de revisión;(ii) comunicarse con el grupo editorial; (iii) asegurarse de que la revisión esté preparada dentro de las fechas acordadas;(iv) enviar la revisión al grupo editorial; (v) hacer llegar los comentarios editoriales a los coautores; y (vi) asegurarse de que las actualizaciones estén preparadas.

La persona de contacto no necesita encabezar la lista de autores, y la elección de la persona de contacto no debe alterar las citas en la revisión. Si una persona de contacto no quiere seguir siendo el responsable de una revisión publicada y otro miembro del equipo de revisión no quiere reemplazarlo, entonces se proporcionarán los detalles de contacto del Coordinador del Grupo de Revisión (RGC, siglas en inglés). La persona de contacto de una revisión no requiere ser mencionada como un autor.

4.2.4 Fechas

4.2.4.1 Evaluada como actualizada

La última fecha en que la revisión se consideró actualizada a menudo coincidirá con la fecha en que los autores enviaron la revisión para ser publicada en la *Base de Datos Cochrane de Revisiones Sistemáticas (CDSR sigla en inglés).*

Ver también

- Los criterios específicos para describir una revisión como actualizada aparecen en el Capítulo 3 (sección 3.2).

4.2.4.2 Fecha de búsqueda

Esta fecha es utilizada para ayudar a determinar si una revisión ha sido actualizada, y para informar sobre la fecha en la que la revisión es considerada como una actualización. No será publicada en la *CDSR.*

Ver también

- Los criterios específicos para precisar la fecha de búsqueda aparecen en el Capítulo 3 (Sección 3.3.3).

- Los métodos de búsqueda se discuten en detalle en el Capítulo 6 (Sección 6.3).

4.2.4.3 Etapa siguiente

Es una fecha sólo para uso interno (no será publicada en la *CDSR*) que indica cuándo estará terminada la revisión (para los protocolos) o la próxima actualización de la revisión (para las revisiones).

Ver también

- Las políticas de actualización de las revisiones se describen en el Capítulo 3 (Sección 3.1).

4.2.4.4 Primera publicación del protocolo

Aquí se anota el volumen de la *CDSR* en el que se publicó el protocolo por primera vez (por ejemplo, Volumen 2, 2004). La fecha no será publicada en el RevMan.

4.2.4.5 Primera publicación de la revisión

Aquí se anota el volumen de la *CDSR* en el que se publicó la revisión completa por primera vez (por ejemplo, Volumen I, 2005). La fecha no será publicada en el RevMan.

4.2.4.6 El volumen de la última cita

Aquí se anota el volumen de la *CDSR* en el que se publicó por primera vez la versión actual de la cita de la revisión (por ejemplo, Volumen I, 2007). La fecha no será publicada en el RevMan.

Ver también

- La cita de las versiones se discute en detalle en el Capítulo 3 (Sección 3.2).

4.2.5 Novedades e Historia

La sección ‘Novedades’ debe describir los cambios en el protocolo o en la revisión desde su última publicación en la *CDSR.* En cada actualización o enmienda de una revisión, debe registrarse al menos un evento de ‘Novedades’ que contenga el tipo de evento, la fecha del cambio y una descripción de lo que se ha cambiado. Esta descripción puede ser, por ejemplo, un breve resumen de la cantidad de información nueva que se agregó a la revisión (por ejemplo, número de estudios, participantes o análisis adicionales) o cualquiera de los cambios importantes hechos a las conclusiones, los resultados o los métodos de la revisión. Las anotaciones de la tabla ‘Novedades’ que no se relacionen con la versión actual de la cita de la revisión deben incluirse en ‘Historia’.

Ver también

- Los eventos de la tabla ‘Novedades’ se discuten en detalle en el capítulo 3 (sección 3.5).

4.3 Resumen

Todas las revisiones completas deben incluir un resumen de 400 palabras o menos. El resumen debe abreviar sin sacrificar contenido importante. Los resúmenes de las revisiones Cochrane se publicarán en MEDLINE y en el Science Citation Index, y estarán disponibles libremente en Internet. Por lo tanto es importante que puedan leerse como documentos aislados.

Ver también

- La guía para el contenido de un resumen se encuentra en el Capítulo 11 (Sección 11.8.).

4.4 Resumen en términos sencillos

El resumen en términos sencillos (antes llamado la 'sinopsis') pretende abreviar la revisión en un estilo franco, que pueda ser comprendido por los usuarios de la sanidad. Los resúmenes en términos sencillos están disponibles libremente en Internet, por lo que a menudo serán leídos como documentos aislados. Los resúmenes en términos sencillos tienen dos partes: un título en términos sencillos (una ratificación del título de la revisión utilizando términos del lenguaje corriente) y un texto de no más de 400 palabras.

Ver también

- La guía para el contenido de una reseña en lenguaje corriente se encuentra en el Capítulo 11 (Sección 11.9).

4.5 Texto Principal

El texto de la revisión debe ser concreto y legible. Aunque no se ha establecido formalmente un número de palabras límite para las revisiones Cochrane, los autores de las revisiones deberán considerar 10.000 palabras como la cifra máxima absoluta, a menos que haya una razón especial para escribir una revisión más extensa. Una revisión debe escribirse de tal manera que una persona que no sea experta en el tema pueda entenderla, de acuerdo con la siguiente declaración, mencionada en el Manual Cochrane (www.cochrane.org/admin/manual.htm):

"El objetivo de las revisiones Cochrane es que su audiencia esté formada por quienes toman decisiones en temas sanitarios. Esto incluye profesionales de la salud, usuarios y legisladores que tengan una comprensión básica de la enfermedad o del problema en cuestión.

Es una parte de la misión y un principio básico de la Colaboración Cochrane promover el acceso a las revisiones sistemáticas sobre el efecto de las intervenciones sanitarias para quienquiera que desee tomar decisiones en temas de salud. Sin embargo, esto no significa que las revisiones Cochrane sean comprensibles para cualquiera sin importar su formación. Esto no es posible, como tampoco que las revisiones Cochrane sean escritas en un único lenguaje comprensible para todas las personas en el mundo.

Las revisiones Cochrane deben ser escritas de tal manera que sean fáciles de leer y comprender para alguien con un conocimiento básico del tema, que no necesite ser experto en el mismo. Aun así, alguna explicación de términos y conceptos puede ser útil, incluso necesaria. Sin embargo, demasiada explicación puede entorpecer la legibilidad de una revisión. La simplicidad y la claridad también son vitales para la legibilidad. La legibilidad de las revisiones Cochrane debe ser comparable a la de un artículo bien escrito en una revista médica general".

El texto de una revisión Cochrane contiene un número fijo de títulos y subtítulos que están incluidos en el RevMan. Un autor podrá agregar más subtítulos en cualquier momento. Se recomiendan algunos subtítulos para que sean usados por todos los autores (están incluidos en el RevMan como opcionales), pero no son obligatorios y deben evitarse si hacen determinadas secciones innecesariamente cortas. Más subtítulos que pueden ser o no relevantes para una revisión en particular también se discuten abajo.

Los siguientes títulos fijos son seguidos de subtítulos fijos y no deben tener texto libre inmediatamente después de ellos: 'Métodos', 'Criterios pata la inclusión de estudios', 'Resultados', y 'Conclusiones de los Autores'.

Antecedentes [fijo, título de nivel 1]

Las preguntas bien formuladas para una revisión se enmarcan en el contexto de un cuerpo de conocimiento ya formado. La experiencia debe dirigir este contexto, ayudar a establecer la razón de la revisión, y explicar por qué son importantes las preguntas que se han hecho. Esto debe ser conciso (por lo general alrededor de una página impresa) y comprensible para los usuarios de la intervención bajo investigación. Se deben citar todas las fuentes de información.

Descripción de la afección [recomendado, título de nivel 2]

La revisión debe comenzar con una breve descripción de la enfermedad considerada y su importancia. Puede incluir información sobre la biología, el diagnóstico, el pronóstico y su importancia en salud pública (incluida la prevalencia o la incidencia).

Descripción de la intervención [recomendado, título de nivel 2]

Una descripción de la(s) intervención(es) experimental(es) debe situarla(s) en un contexto de intervenciones corrientes o alternativas. El rol de la(s) intervención(es) comparativa(s) en la práctica corriente debe quedar claro. Para los medicamentos se debe presentar información farmacológica básica cuando esté disponible. Esta información puede incluir dosis, metabolismo, efectos selectivos, vida media, tiempo de acción, y cualquiera de las interacciones conocidas con otras drogas. Para intervenciones más complejas se debe aportar información de los principales componentes.

Cómo podría funcionar la intervención [recomendado, título de nivel 2]

Esta sección describirá las razones teóricas por las que las intervenciones revisadas pueden tener un impacto en los destinatarios potenciales, por ejemplo, al relacionar una intervención farmacológica con la biología de la enfermedad. Los autores pueden referirse a un conjunto de evidencia empírica, tal como intervenciones similares que tengan un impacto, o intervenciones idénticas que tengan un impacto en otras poblaciones. Los autores pueden referirse también a un conjunto de literatura que justifique la posibilidad de la eficacia.

Por qué es importante hacer esta revisión [recomendado, título de nivel 2]

Los antecedentes deberían establecer claramente la razón de la revisión y deberían explicar por qué las preguntas que se hacen son importantes. También podrían mencionar por qué se eligió esta revisión, y cómo se la podría relacionar con una revisión más amplia de un problema general. Si esta versión de la revisión es una actualización de una anterior es útil establecerlo por escrito, por ejemplo, “Esta es una actualización de una revisión Cochrane publicada inicialmente en el AÑO, y previamente actualizada en el AÑO”. Puede ser complementada con una descripción breve de los principales hallazgos de las versiones precedentes, con una mención de cualquiera de las razones específicas que se dieron para actualizar la revisión.

Objetivos [fijo, título de nivel 1]

Debe comenzar con una definición precisa del objetivo primario de la revisión, preferentemente con una frase sencilla. En lo posible la forma debería ser “Evaluar los efectos de [intervención o comparación] para [problema de salud] para/en [grupos de personas, enfermedad o problema, y lugar si se ha definido]”. Esto podría ir seguido de una serie de objetivos específicos relacionados con los diferentes grupos participantes, diferentes comparaciones de intervenciones o diferentes medidas de desenlace. No es necesario plantear hipótesis específicas.

Métodos [fijo, título de nivel 1]

La sección Métodos de un protocolo debería escribirse siempre en tiempo futuro. Como las revisiones Cochrane son actualizadas a medida que se acumula nueva evidencia, los métodos definidos en el protocolo deberían escribirse como si un gran número de estudios fueran a ser identificados adecuadamente para permitir el logro de los objetivos (aunque esto se sepa, no es el caso en el momento de escribir).
La sección Métodos de una revisión debería estar escrita en tiempo pasado, y debería describir lo que se hizo para obtener *los resultados y las conclusiones de la presente revisión*. Se anima a los autores de la revisión a que citen su propio protocolo para aclarar que sí lo hubo. A menudo en una revisión no se pueden implementar todos los métodos diseñados en el protocolo, generalmente porque la evidencia es insuficiente. En tales casos se recomienda que los métodos que no fueron implementados se incluyan en la sección titulada ‘Diferencias entre el protocolo y la revisión’ (ver abajo), de tal manera que puedan servir como protocolo para futuras actualizaciones de la revisión.

Criterios para considerar estudios para esta revisión [fijo, título de nivel 2]

Tipos de estudios [fijo, título de nivel 3]

Los diseños de estudios elegibles deben definirse aquí, junto con cualquiera de los criterios de inclusión basados en la conducta de los estudios o su riesgo de sesgo. Por ejemplo, 'Todas la comparaciones controladas aleatorizadas' o 'Todos los ensayos controlados aleatorizados con evaluación ciega de resultados'. Se debe justificar la exclusión de determinados tipos de estudios aleatorizados (por ejemplo, ensayos cruzados).

Ver también

- Los criterios de elegibilidad para los tipos de diseño de estudios se discuten en el Capítulo 5 (Sección 5.5).

Tipos de participantes [fijo, título de nivel 3]

Las enfermedades o situaciones de interés deberían describirse aquí, incluidas cualquiera de las restricciones como diagnósticos, grupos de edad y entornos. Los análisis de los subgrupos no deben incluirse aquí (ver 'Análisis de subgrupos e investigación de heterogeneidad' en 'Métodos').

Ver también

- Los criterios de elegibilidad para los tipos de participantes se discuten en el Capítulo 5 (Sección 5.2).

Tipos de intervenciones [fijo, título de nivel 3]

Las intervenciones experimentales y comparativas deben definirse aquí, en subtítulos separados si es el caso. Se debería aclarar cuáles son las comparaciones de interés. Se deben establecer las restricciones en dosis, frecuencia, intensidad o duración. Los subgrupos de análisis no se deben incluir aquí (ver 'Análisis e investigación de subgrupos de heterogeneidad', en 'Métodos').

Ver también

- Los criterios de elegibilidad para los tipos de intervenciones se discuten en el Capítulo 5 (Sección 5.3).

Tipos de medidas de desenlace [fijo, título de nivel 3]

Notar que las medidas de desenlace no siempre forman parte de los criterios para la inclusión de estudios en una revisión. Si no forman parte, esto debe quedar claro. Las medidas de desenlace que interesen deben incluirse aquí, formen parte o no de los criterios de elegibilidad.

Ver también

- Los tipos de desenlaces se discuten en el Capítulo 5 (Sección 5.4).
- La importancia de enfocar los desenlaces relevantes para el paciente se discute más en el Capítulo 11 (Sección 11.5.2); ver también una amplia discusión de los desenlaces informados por el paciente en el Capítulo 17.

Desenlaces principales [*recomendado, título de nivel 4]*

Los principales desenlaces de la revisión normalmente deberían reflejar al menos un beneficio potencial y por lo menos un área de daño potencial, y deben ser tan pocos como sea posible. Normalmente se espera que la revisión sea capaz de analizar estos desenlaces

si se identifican los estudios elegibles, y que las conclusiones de la revisión deban basarse principalmente en los efectos que tienen las intervenciones en estos desenlaces.

Desenlaces secundarios *[recomendado, título de nivel 4]*

Los desenlaces no principales deben incluirse aquí. El número total de desenlaces considerados debe mantenerse tan reducido como sea posible.

Los siguientes títulos *opcionales* (nivel 4) pueden ser útiles, como un complemento o un reemplazo de los títulos anteriores:

Desenlaces importantes para la tabla 'Resumen de los resultados'

Momento de la evaluación de desenlaces

Desenlaces adversos

Datos económicos

Métodos de búsqueda para la identificación de estudios [fijo, título de nivel 2]

Los métodos utilizados para identificar estudios deberán resumirse. Se recomiendan los siguientes títulos. Antes de empezar a desarrollar esta sección los autores deben contactar con su Grupo Cochrane de Revisión (CRG, sigla en inglés) para obtener ayuda.

Ver también
- Los métodos de búsqueda se discuten en detalle en el Capítulo 6 (Secciones 6.3).

Búsquedas electrónicas [recomendado, título de nivel 3]

Las bases de datos bibliográficas consultadas, las fechas y períodos consultados y cualquiera de las limitaciones, tales como el lenguaje, deben explicitarse. Las estrategias completas de búsqueda para cada base de datos deberían listarse en un anexo de la revisión. Si un CRG ha desarrollado un registro especializado de estudios y éste es consultado para la revisión, debe mencionarse una descripción estándar de este registro. Deberá incluirse información sobre cuándo y cómo fue consultado más recientemente el registro especializado para la versión actual de la revisión; asimismo, deben anotarse los términos de búsqueda utilizados.

Ver también
- Las estrategias de búsqueda se discuten en detalle en el Capítulo 6 (Sección 6.4).

Buscando otras fuentes [recomendado, título de nivel 3]

Enumeran fuentes de literatura gris, como informes internos y memorias de conferencias. Debe anotarse si se ha hecho una búsqueda manual de revistas específica para la revisión, pero la búsqueda manual hecha por los autores para ayudar a construir el registro especializado del CRG no debe anotarse porque esto está incluido en la descripción estandarizada del registro. Enumere las personas (investigadores o especialistas temáticos) y las organizaciones que fueron contactadas. Enumere cualquiera de las otras fuentes utilizadas, que pueden incluir, por ejemplo, listas de referencia, páginas de Internet, o las colecciones personales de artículos.

Los siguientes títulos opcionales pueden utilizarse, bien en lugar de 'Buscando otras fuentes' (en cuyo caso serían títulos de nivel 3) o como subtítulos (nivel 4)

Literatura gris

Búsqueda manual

Listas de referencia

Correspondencia

Ver también
- Otras fuentes de búsqueda se discuten en el Capítulo 6 (Sección 6.2).

Extracción de datos y análisis [fijo, título de nivel 2]

Aquí se describirían los métodos para la extracción de datos y para el análisis.

Selección de estudios [recomendado, título de nivel 3]

Anotar aquí el método utilizado para aplicar los criterios de selección. Debería dejarse claro si han sido aplicados independientemente por más de un autor, así como la manera en que se resolvió cualquiera de los desacuerdos.

Ver también
- La selección de estudios se discute en el Capítulo 7 (Sección 7.2).

Extracción y manejo de los datos [recomendado, título de nivel 3]

Deberá anotarse el método utilizado para extraer u obtener los datos de informes publicados o de los investigadores originales (por ejemplo, utilizando un formulario de extracción de datos).

Deberá mencionarse si los datos fueron extraídos independientemente por más de un autor, así como la manera en que se resolvió cualquiera de los desacuerdos. Si es el caso, los métodos para procesar los datos en preparación para el análisis deberán describirse.

Ver también
- La recolección de los datos se discute en el Capítulo 7, e incluye qué datos se deben recolectar (Sección 7.3), fuentes de datos (Sección 7.4), formas de recolección de datos (Sección 7.5) y la extracción de datos de los informes (Sección 7.6).

Evaluación del riesgo de sesgo en los estudios incluidos [recomendado, título de nivel 3]

Se deberá mencionar el método utilizado para evaluar el riesgo de sesgo (o calidad metodológica).

Deberá establecerse si los métodos fueron aplicados independientemente por más de un autor, así como la manera en que se resolvió cualquiera de los desacuerdos que se presentaron. La(s) herramienta(s) utilizadas deberán describirse o referenciarse, indicando además cómo se incorporaron los resultados a la interpretación de los resultados de efecto de las intervencionse.

Ver también
- La herramienta recomendada para hacerlo se describe en el Capítulo 8 (Sección 8.5).

Medidas del efecto del tratamiento [recomendado, título de nivel 3]

Se deben mencionar las medidas de efecto elegidas. Por ejemplo, odds ratio (OR, sigla en inglés), razón de riesgos (RR) o diferencia de riesgos (RD) para los datos dicotómicos; diferencia de medias (MD) o diferencia estandarizada de medias (SMD) para datos continuos. Los siguientes títulos opcionales pueden utilizarse, bien en lugar de 'Medidas del efecto del tratamiento' (en cuyo caso serían títulos de nivel 3) o como subtítulos (nivel 4):

Datos dicotómicos

Datos continuos

Datos' tiempo al suceso'

Ver también
- Los tipos de medidas de datos y efectos se discuten en el Capítulo 9 (Sección 9.2).

Aspectos relacionados con la unidad de análisis [recomendado, título de nivel 3]

Deberán describirse temas especiales en el análisis de estudios con diseños no estándar, tales como ensayos cruzados y ensayos aleatorizados por conglomerados. Alternativamente, pueden usarse títulos *opcionales* (nivel 3) específicos para los tipos de estudio, como:

Ensayos aleatorizados por conglomerados

Ensayos cruzados

Estudios con múltiples grupos de tratamiento

Ver también
- Los temas de la unidad de análisis se discuten en el Capítulo 9 (Sección 9.3).
- Algunos diseños no estandarizados se discuten en detalle en el Capítulo 16, incluidos los ensayos aleatorizados por conglomerados (Sección 16.3), ensayos cruzados (Sección 16.4), y estudios con múltiples grupos de tratamiento (Sección 16.5). Los estudios no aleatorizados se discuten en el Capítulo 13.

El manejo de los datos no disponibles [recomendado, título de nivel 3]

Deberán describirse las estrategias para el manejo de los datos no disponibles. Esto incluiría principalmente a participantes perdidos por abandono (y si se llevará a cabo un análisis de la intención de tratar) y a estadísticos no disponibles (tales como desviaciones estándar y coeficientes de correlación).

Ver también
- Los aspectos relacionados con los datos no disponibles se discutirán en el Capítulo 16 (Secciones 16.1) y los aspectos de intención de tratar en el Capítulo 16 (Sección 16.2).

Evaluación de heterogeneidad [recomendado, título de nivel 3]

Se deberían describir los enfoques para manejar la heterogeneidad clínica, así como la manera en que los autores determinarán si se considera apropiado un metaanálisis. Los métodos para identificar la heterogeneidad estadística se deberán definir (por ejemplo visualmente, utilizando I^2, haciendo una prueba del chi cuadrado).

Ver también
- La evaluación de la heterogeneidad se discutirá en el Capítulo 9 (Sección 9.5).

Evaluación de sesgos de reporte [recomendado, título de nivel 3]

Esta sección deberá describir cómo se maneja el sesgo de publicación y otros sesgos de reporte (por ejemplo, gráficos de embudo, pruebas estadísticas, imputación). Los autores deberán recordar que la asimetria de un gráfico de embudo no está necesariamente causada por el sesgo de publicación (y que los sesgos de publicación no necesariamente causan asimetría en un gráfico de embudo).

Ver también
- Los sesgos de reporte se discuten en el Capítulo 10.

Síntesis de los datos [recomendado, título de nivel 3]

Se debería establecer la elección del método del metanálisis, incluyendo si se ha utilizado un modelo de efecto fijo o de efectos aleatorios . Si no se han realizado metanálisis se deberán describir los enfoques sistemáticos utilizados para sintetizar los hallazgos de los múltiples estudios.

Ver también
- Metanálisis y la síntesis de los datos se discuten en el Capítulo 9 (Sección 9.4).

Análisis de subgrupo e investigación de heterogeneidad [recomendado, título de nivel 3]

Todos los análisis de los subgrupos planificados deberán enumerarse (o las variables independientes de metaregresión). Cualquier otro método utilizado para investigar la heterogeneidad de los efectos deberá describirse.

Ver también
- La investigación de la heterogeneidad se discute en el Capítulo 9 (Sección 9.6).

Análisis de sensibilidad [recomendado, título de nivel 3]

Deberán describirse los análisis que pretenden determinar si las conclusiones refuerzan las decisiones tomadas durante el proceso de revisión, tales como la inclusión/exclusión de determinados estudios de un metanálisis, la imputación por falta de datos o la elección de un método de análisis.

Ver también
- El análisis de sensibilidad se discute en el Capítulo 9 (Sección 9.7).

Estos otros títulos opcionales (nivel 3) para la sección Métodos pueden ser útiles:

Resumen de la tabla de hallazgos [recomendado, título de nivel 3]

Esta sección debería describir los métodos utilizados para preparar cualquier "Resumen de la tabla de hallazgos". Debería incluir información acerca de: (i) qué poblaciones (incluyendo la especificación de las poblaciones de bajo, medio o alto riesgo), las intervenciones y comparaciones que se analizan en una o más tablas de "Resumen de hallazgos" y por qué; (ii) las fuentes de cualquier información externa utilizada en la columna de "Riesgos asumidos"; (iii) un comentario breve de que se utilizan los criterios GRADE para valorar la calidad de la evidencia; i (iV) cualquier desviación de los métodos estándar descritos en el Capítulo 11 y Capítulo 12, con una justificación para estas modificaciones. Los principales desenlaces de la revisión, por ej. aquellos que van a ser incluidos en la tabla de "Resumen de hallazgos" deberían quedar listados en la sección "Tipos de medidas de resultado".

Ver también

- Las tablas de "Resumen de hallazgos" se discuten en el Capítulo 11 (Sección 11.5) mientras que el sitema GRADE se describe en el Capítulo 12 (Sección 12.2) y el archivo de ayuda en el software GRADEpro.

Nota: este título no se incluye por defecto en el software Revman pero puede ser incorporado por el autor de la revisión.

El siguiente título *adicional* (nivel 3) de la sección de Métodos puede ser útil:

Aspectos económicos

Métodos para actualizaciones futuras

Los autores que busquen analizar aspectos económicos de las intervenciones en una revisión deberán considerarlo desde las primeras etapas del desarrollo de un protocolo.

Ver también
- Los aspectos económicos se discuten en el Capítulo 15.
- Los aspectos de actualización de las revisiones se discuten en el Capítulo 3.

Resultados [fijo, título de nivel 1]

Descripción de los estudios [fijo, título de nivel 2]

Resultados de la búsqueda [recomendado, título de nivel 3]

Las secciones de resultados deberían comenzar con un resumen de los resultados de la búsqueda (por ejemplo, cuantas referencias se recuperaron mediante las búsquedas electrónicas y cuantas se consideran potencialmente legibles después de cribarlas). Se recomienda ámpliamente a los autores de la revisiones que incluan un diagrama del flujo de estudios tal como está formulado en los criterios PRISMA (Liberati 2009). Estos diagramas de flujo se pueden crear con el RevMan 5.1.

Ver también

- La presentación de los diagramas de flujo se discuten en el Capítulo 11 (Sección 11.2.1)

Estudios incluidos [recomendado, título de nivel 3]

Es importante que se establezca claramente el número de estudios incluidos. Esta sección deberá incluir un breve resumen de la información contenida en la tabla 'Características de los estudios incluidos'. Se debe incluir una referencia explícita a esta tabla. Se deberán describir las características clave de los estudios incluidos, que contengan los participantes en el estudio, lugar (por ejemplo, país), entorno (si es importante), intervenciones, comparaciones y medidas de desenlace de los estudios incluidos y cualquiera de las diferencias importantes entre los estudios. Debería de mencionarse aquí el sexo y el rango de edad de los participantes excepto cuando su naturaleza sea obvia (por ejemplo, si todas las participantes están embarazadas). Deberían aportarse los detalles importantes de las intervenciones específicas utilizadas (por ejemplo para radioterapia se debería anotar la dosis total, el número de sesiones y el tipo de radiación utilizada; para los medicamentos se debe resumir su preparación, vía de administración, dosis y frecuencia). Los autores deberían anotar otras características de los estudios que ellos consideren importante que conozcan los lectores de la revisión.
Los siguientes subtítulos opcionales (nivel 4) pueden ser útiles:

Diseño

Tamaño de la muestra

Ámbito

Participantes

Intervenciones

Desenlaces de interés

Ver También
- La tabla 'Características de los estudios incluidos' (Sección 4.6.1).

Estudios excluidos [recomendado, título de nivel 3]

Aquí habría que referirse a la información contenida en la tabla 'Características de los estudios excluidos'. Se deberá incluir una referencia explícita a esta tabla, y anexar un breve resumen de por qué fueron excluidos estos estudios de la revisión.

Ver también
- Tabla 'Características de los estudios excluidos' (Sección 4.6.3).

Los siguientes títulos opcionales (nivel 3) pueden utilizarse en la sección 'Descripción de los estudios':

Estudios en marcha

Estudios pendientes de clasificación

Estudios nuevos hallados en esta actualización

Riesgo de sesgo en los estudios incluidos [fijo, título de nivel 2]

Aquí debería resumirse el riesgo de sesgo general en los resultados de los estudios incluidos, su variabilidad a través de los estudios y cualquier defecto importante en los

estudios individuales. Los criterios que fueron utilizados para evaluar el riesgo de sesgo deberán ser descritos o referenciados en 'Métodos' y no aquí. La forma en que se clasificó cada estudio en uno de los criterios determinados deberá informarse en una tabla de 'riesgo de sesgo' y no describirse detalladamente dentro del texto, el cual debe ser un resumen conciso.

Ver también

- La presentación de las evaluaciones del 'riesgo de sesgo' se incluyen en el Capítulo 8 (Sección 8.6).

En las revisiones extensas los aspectos de la evaluación del riesgo de sesgo deben resumirse para los principales resultados bajo los siguientes títulos.

Asignación [recomendado, título de nivel 3]

Aquí debe resumirse brevemente cómo se generaron las secuencias de asignación y los intentos de ocultar la asignación de la intervención, junto con cualquier juicio de valor relacionado con el riesgo de sesgo que pueda derivarse de los métodos utilizados.

Cegamiento [recomendado, título de nivel 3]

Aquí se deberá informar con un breve resumen sobre quién fue cegado o enmascarado durante el curso o el análisis de los estudios. Las implicaciones del cegamiento de la evaluación de los resultados pueden ser diferentes para resultados diversos, así que deben ser valorados por separado. Deben resumirse los juicios de valor relacionados con el riesgo de sesgo asociado con el cegamiento.

Datos incompletos de los resultados [recomendado, nivel 3]

La integridad de los datos deberá resumirse brevemente aquí para cada uno de los principales resultados. Deberán informarse las inquietudes de los autores de la revisión por la exclusión de participantes y por abandonos excesivos.

Informe selectivo [recomendado, título de nivel 3]

Las inquietudes sobre la disponibilidad selectiva de los datos deberán resumirse brevemente aquí, incluida la evidencia del informe selectivo de los desenlaces, tiempos, subgrupos o análisis.

Otras fuentes potenciales de sesgo [recomendado, título de nivel 3]

Cualquier otra de las inquietudes potenciales debería resumirse aquí.

Efectos de las intervenciones [fijo, título de nivel 2]

Éste deberá ser un resumen de los principales hallazgos de los efectos de las intervenciones estudiadas en la revisión. La sección deberá mencionar directamente los objetivos de la revisión en vez de enumerar los resultados de los estudios incluidos uno por uno. Los desenlaces de los estudios individuales y cualquier resumen estadístico de éstos deberán incluirse en las tablas 'Datos y análisis'. Los desenlaces deberán ser mencionados normalmente en el orden en el que se enumeran en 'Tipos de medidas de desenlace'. Se recomienda el uso de los subtítulos si facilitan la comprensión (por ejemplo, para cada grupo participante, comparación o medida de desenlace diferentes, si la revisión analiza más de uno). Deberá informarse sobre cualquiera de los análisis de sensibilidad que se haya realizado.

Los autores no deberán hacer inferencias en esta sección. Un error común que debe evitarse (tanto en la descripción de los resultados como en la extracción de las conclusiones) es confundir 'no hay evidencia de un efecto' con 'hay evidencia de no efecto". Cuando la evidencia no es concluyente es un error declarar que una intervención 'no tiene efecto' o que 'no es diferente' de la intervención de control. En este caso es más seguro informar el dato con un intervalo de confianza, que es compatible tanto con una reducción como con un aumento en el resultado.

Ver también

- La presentación de los resultados se encuentra en el Capítulo 11 (Sección 11.7).
- La interpretación de los resultados numéricos se discute en el Capítulo 12 (Secciones 12.4, 12.5 y 12.6).

Discusión [fijo, título de nivel 1]

Una discusión estructurada puede ayudar en la consideración de las implicaciones de la revisión (Docherty 1999).

Ver también

- La interpretación de los resultados se discute en el Capítulo 12.

Resumen de los principales resultados [recomendado, título de nivel 2]

Resumir los hallazgos más importantes (sin repetir la sección 'Efectos de las intervenciones') y las incertidumbres más destacadas, sopesando los beneficios importantes junto con los daños principales. Referirse explícitamente a cualquiera de las tablas 'Resumen de los resultados'.

Exhaustividad y aplicabilidad de la evidencia [recomendado, título de nivel 2]

Describir la relevancia de la evidencia respecto la pregunta de la revisión. Esto debería llevar a un juicio general de la validez externa de la revisión. ¿Los estudios identificados son suficientes para alcanzar todos los objetivos de la revisión? ¿Han sido investigados todos los tipos relevantes de participantes, intervenciones y resultados? Los comentarios de cómo los resultados de la revisión encajan dentro del contexto de la práctica corriente podrían incluirse aquí, aunque los autores deberán tener en mente que la práctica corriente puede variar internacionalmente.

Calidad de la evidencia [recomendado, título de nivel 2]

¿El conjunto de la evidencia identificada permite una conclusión robusta acerca del objetivo(s) de la revisión? Hay que resumir la evidencia que se ha incluido (número de estudios, número de participantes), detallar las limitaciones metodológicas de los estudios, y reiterar la consistencia o inconsistencia de los resultados. Todo ello debería llevar a una descripción del juicio global de la calidad de la evidencia que contribuye a los resultados de las revisión. Las razones o justificaciones para incrementar o degradar la calidad de la evidencia en la tabla "Resumen de hallazgos" (tal como se describe en las notas de aquella tabla) también deberían ser descritas.

Sesgos potenciales en el proceso de revisión [recomendado, título de nivel 2]

Se deberían establecer las fortalezas y limitaciones de la revisión en lo que respecta a la prevención de sesgo. Estos factores pueden estar dentro o fuera del control de los autores de la revisión. La discusión podría incluir la posibilidad de que todos los estudios relevantes fueran identificados, de que todos los datos relevantes pudieran ser obtenidos, o de que los

métodos empleados podrían haber introducido algún sesgo (por ejemplo, búsqueda, selección de estudios, recolección de datos, análisis).

Acuerdos y desacuerdos con otros estudios o revisiones [recomendado, título de nivel 2]

Deberían incluirse aquí los comentarios sobre cómo los estudios incluidos encajan dentro del contexto de otra evidencia, diciendo claramente si la otra evidencia fue revisada sistemáticamente.

Conclusiones de los autores [fijo, título de nivel 1]

El propósito principal de la revisión debe ser presentar información más que ofrecer consejo. Las conclusiones de los autores están divididas en dos secciones:

Implicaciones para la práctica [fijo, título de nivel 2]

Las implicaciones para la práctica deben ser tan útiles y claras como sea posible. No deberían ir más allá de la evidencia que se revisó y estar justificadas por los datos presentados en la revisión. 'No hay evidencia de efecto' no se debe confundir con 'hay evidencia de no efecto'.

Implicaciones para la investigación [fijo, título de nivel 2]

Esta sección de las revisiones Cochrane es utilizada cada vez más por personas que toman decisiones sobre investigaciones futuras, y los autores deberían tratar de escribir algo que pueda ser útil para este propósito. Como en 'Implicaciones para la práctica', el contenido debería basarse en la evidencia disponible y debería evitar el uso de información que no esté incluida o no se haya discutido en la revisión.

Al preparar esta sección los autores deberán considerar los diferentes aspectos de la investigación, quizá utilizando como marco los tipos de estudio, los participantes, las intervenciones y los desenlaces. Las implicaciones de *cómo* debería hacerse e informarse la investigación deberían distinguirse de *qué* investigación futura debería hacerse. Por ejemplo, la necesidad de ensayos aleatorizados en vez de otros tipos de estudios para mejores descripciones de temas particulares de la revisión, o para la obtención de resultados específicos, deberá distinguirse de la superfluidad de más comparaciones con placebo si hay un tratamiento activo efectivo y apropiado, o de la necesidad de comparaciones de intervenciones específicas, o de investigación en tipos específicos de personas.

Es importante que esta sección sea tan clara y explicita como sea posible. Expresiones generales que contienen poca o ninguna información específica, como 'debe realizarse mejor investigación en el futuro', o 'se necesita más investigación' son poco útiles para las personas que toman decisiones, y deberían evitarse.

Ver también

- La guía para presentar las conclusiones se proporciona en el Capítulo 12 (Sección 12.7).

Agradecimientos [fijo, título de nivel 1]

Esta sección deberá ser utilizada para mencionar a las personas o instituciones a quienes los autores quieran dar las gracias, incluidas personas que no estén mencionadas como autores. Este reconocimiento incluiría a cualquiera de los autores previos de la revisión Cochrane, o fuentes previas de apoyo a la revisión, y podría incluir las contribuciones del

grupo editorial del CRG. Se debe obtener el permiso de las personas para mencionarlas en este capítulo.

Contribuciones de los autores [fijo, título de nivel 1]

Las contribuciones de los coautores actuales del protocolo o de la revisión deberían describirse en esta sección. Debería identificarse un autor como el garante de la revisión. Todos los autores deben comentar y ponerse de acuerdo en la descripción de sus respectivas contribuciones antes de que la revisión sea enviada para publicación en la *CDSR*. Cuando la revisión se actualiza esta sección debería revisarse a fondo para asegurarse de que está correcta y actualizada.

Las siguientes potenciales contribuciones se han adaptado de Yank y cols. (Yank 1999).
Esta es una sugerencia y la sección debería describir qué hicieron los autores, más que intentar identificar dentro de cuál de estas categorías se puede incluir la contribución de alguien. Es ideal que los autores describan su contribución con sus propias palabras:

- Concebir la revisión;
- Diseñar la revisión;
- Coordinar la revisión;
- Obtener los datos para la revisión;
  - Diseñar las estrategias de búsqueda;
  - Desarrollar búsquedas;
  - Filtrar los resultados de la búsqueda;
  - Organizar la recuperación de los artículos;
  - Filtrar los artículos recuperados utilizando los criterios de elegibilidad;
  - Valorar la calidad de los artículos;
  - Extraer datos de los artículos;
  - Escribir a los autores de los artículos pidiendo información adicional;
  - Aportar datos adicionales sobre los artículos;
  - Obtener y filtrar datos de estudios no publicados;
- Manejo de datos para la revisión;
  - Entrar datos al RevMan;
- Análisis de los datos;
- Interpretación de los datos;
  - Aportar una perspectiva metodológica;
  - Aportar una perspectiva clínica;
  - Aportar una perspectiva de gestión;
  - Aportar una perspectiva de los consumidores;
- Escribir la revisión (o el protocolo);
- Aportar consejos generales sobre la revisión;
- Asegurar fondos para la revisión;
- Realizar el trabajo previo que fue la base de la revisión actual.

Declaraciones de interés [fijo, título de nivel 1]

Los autores deberán informar sobre cualquiera de las afiliaciones, presentes o pasadas u otra vinculación, con cualquier organización o entidad que tenga interés en la revisión que pudiera conducir a un conflicto de interés real o aparente. Situaciones que pudieran ser percibidas por otros como capaces de influir en los juicios del autor de una revisión, incluidos los conflictos personales, políticos, académicos, y otros posibles, tanto como conflictos financieros. Los autores deben declarar si han estado vinculados a algún estudio incluido en la revisión.

Ver también

- Un resumen de las directrices de la Colaboración sobre los conflictos de interés aparece en el Capítulo 2 (Sección 2.6).

Los conflictos de interés causan gran preocupación y deben evitarse, pero sobretodo debe informarse si existe alguno. Debe informarse de cualquier interés secundario (como los conflictos personales) que pueda influir demasiado en los juicios realizados en una revisión (relacionados por ejemplo con la inclusión o exclusión de estudios, evaluaciones de la validez de los estudios incluidos, o la interpretación de los resultados).

Si no hay conflictos de interés conocidos debe mencionarse explícitamente, por ejemplo escribiendo 'Ninguno conocido'.

Diferencias entre el protocolo y la revisión [fijo, título de nivel 1]

En ocasiones es necesario utilizar métodos diferentes de aquellos descritos en el protocolo original. Esto puede deberse a:

- Los métodos para tratar tema particular no se han especificado en el protocolo;
- No se pudieron aplicar los métodos incluidos en el protocolo (por ejemplo, debido a datos insuficientes o a falta de información requerida para implementar los métodos);
- Los métodos se cambiaron porque se descubrió una alternativa preferible.

Es aceptable introducir algunos cambios de los métodos del protocolo a la revisión, pero deben ser ampliamente descritos en esta sección. La sección aporta un resumen de los principales cambios de los métodos de la revisión a lo largo del tiempo y debería utilizarse para:

- Señalar cualquiera de los métodos que se eligieron después de haber publicado el protocolo original (por ejemplo, agregar o cambiar resultados; agregar 'Riesgo de sesgo' o tablas de 'Resumen de los resultados';

- Resumir los métodos del protocolo que no fueron implementados en la revisión actual (por ejemplo, porque la revisión identificó estudios no elegibles, o porque los estudios no se englobaban en ningún subgrupo determinado previamente definido;

- Explicar cualquiera de los cambios realizados en los métodos de los protocolos en la revisión, decir cuándo fueron realizados y dar la justificación para esos cambios. Tales cambios no deberían ser guiados por los hallazgos en los efectos de las intervenciones. Tener en cuenta los posibles efectos en las conclusiones de la revisión de algunos cambios en los métodos, y considerar los análisis de sensibilidad para evaluarlo.

Notas publicadas [fijo, título de nivel 1]

Las notas publicadas aparecerán en la revisión en la *CDSR*. Pueden incluir notas editoriales y comentarios del CRG, por ejemplo, donde aspectos destacados por los editores se consideran útiles para ser publicados a lo largo de la revisión. Se debe especificar el autor o la fuente de estos comentarios (por ejemplo, un autor o un editor).

Las notas publicadas deben estar completadas en todos los protocolos y las revisiones retiradas, dando las razones para su retiro. Para los protocolos y las revisiones retiradas sólo se publicará información de citas básicas, fuentes de apoyo y notas publicadas.

4.6 Tablas

4.6.1 Características de los estudios incluidos

La tabla 'Características de los estudios incluidos' tiene cinco entradas para cada estudio: Métodos, Participantes, Intervenciones, Desenlaces y Notas. Pueden especificarse hasta tres entradas más para temas no cubiertos adecuadamente por estas categorías, por ejemplo, aportar información sobre la duración del seguimiento, fuentes de financiación, o indicaciones de calidad de estudios que difícilmente lleven directamente a un riesgo de sesgo (ver Sección 4.6.2 para información incluida sobre riesgo de sesgo). En la tabla se pueden usar códigos o abreviaturas para facilitar una presentación clara y breve de múltiples partes de información dentro de una entrada; por ejemplo, los autores podrían incluir país, lugar, edad y sexo bajo la entrada 'Participantes'. Deberán escribirse notas a pie de página para explicar cualquiera de los códigos o abreviaturas utilizados (que serán publicados en la *CDSR*).

Ver también

- La guía detallada de 'Las características de los estudios incluidos' se encuentra en el Capítulo 11 (Sección 11.2).

4.6.2 Riesgo de sesgo

Una tabla de 'Riesgo de sesgo' es una extensión opcional, altamente recomendada, de la tabla 'Características de los estudios incluidos'. La tabla estándar 'Riesgo de sesgo' incluye evaluaciones de la generación de secuencia, ocultación de secuencia de asignación, cegamiento, datos incompletos de desenlaces, informe selectivo de desenlaces y 'otros asuntos'. Para cada punto la tabla recoge una descripción de lo que se ha informado que pasó en el estudio y un juicio subjetivo en relación con la protección de sesgo ('Si' para un bajo riesgo de sesgo, 'No' para un alto riesgo de sesgo; 'Poco claro' para lo demás).

Ver también

- Las tablas 'Riesgo de sesgo' se discuten en el Capítulo 8 (Sección 8.6).

4.6.3 Características de los estudios excluidos

Deberán enumerarse ciertos estudios que parecían cumplir los criterios de elegibilidad pero que fueron excluidos, y dar la razón de su exclusión (por ejemplo, intervención comparativa inapropiada). La explicación debería ser breve y usualmente basta con una sola razón para su exclusión.

Ver también

- La selección de los estudios que deben aparecer como excluidos se discute en el Capítulo 7 (Sección 7.2.5).

4.6.4 Características de los estudios pendientes de clasificación

La tabla 'Características de los estudios pendientes de clasificación' (inicialmente 'Estudios pendientes de evaluación') tiene la misma estructura de la tabla 'Características de los estudios incluidos'. Deberá utilizarse para dos categorías de estudios:

- Los estudios en los que no se puede tomar una decisión de incluirlos o excluirlos porque no hay suficiente información disponible en el momento. Se deben hacer todos los esfuerzos razonables para obtener la información antes de dejar de incluirlos en la publicación de la revisión, pero ésta no deberá demorarse excesivamente por esperar esta información, especialmente si la inclusión o exclusión del estudio no tiene impacto en las conclusiones de la revisión. Cuando la información no esté disponible para el acceso a una tabla deberá insertarse el término 'No conocida'.

- Los estudios que han sido identificados pero que están esperando una actualización de la revisión. Es especialmente apropiado mencionar los estudios que tienen un impacto potencial en las conclusiones de la revisión, o estudios que reciben mucha publicidad en el intervalo entre las actualizaciones. Por lo tanto puede hacerse una revisión enmendada con tales estudios resumidos en esta tabla. La actualización completa con todos estos estudios ya incluidos debería terminarse tan pronto como fuera posible. Cuando la información no esté disponible para añadirla a la tabla sería apropiado insertar la expresión 'Aún no evaluada' o 'No conocida'.

4.6.5 Características de los estudios en marcha

Las tablas 'Características de estudios en marcha' tienen ocho entradas para cada estudio: Nombre del estudio, Métodos, Participantes, Intervenciones, Desenlaces, Fecha de comienzo, Información de Contacto y Notas. Los contenidos de estos entradas deberán ser comparables a los de la tabla 'Características de los estudios incluidos'. Las notas a pie de página deberán utilizarse para explicar cualquiera de las abreviaturas empleadas en la tabla (que serán publicadas en la *CDSR*).

4.6.6 Resumen de los hallazgos

Una tabla 'Resumen de los hallazgos' es un medio opcional, aunque altamente recomendado, de presentar los hallazgos de los resultados más importantes, haya o no evidencia disponible de ellos. Una tabla de 'Resumen de hallazgos' incluye, cuando sea necesario, un resumen de toda la evidencia; el riesgo de las personas que reciben intervenciones experimentales y de control; cálculos del efecto relativo (por ejemplo, razón de riesgo u odds ratio); una descripción de la calidad del conjunto de la evidencia; comentarios; y notas de pie de página. A la evaluación del conjunto de la evidencia debe seguir el sistema GRADE, que combina consideraciones sobre el riesgo de sesgo, la confianza, la heterogeneidad, la precisión y la publicación de sesgo.

Ver también
- Las tablas completas de especificación y discusión del 'Resumen de hallazgos' se encuentran en el Capítulo 11 (Sección 11.5);
- El sistema GRADE se revisa en el Capítulo 12 (Sección 12.2).

4.6.7 Tablas adicionales

Se utilizan tablas adicionales para dar información que no puede aparecer adecuadamente publicada en el texto o en las tablas fijas. Algunos ejemplos incluyen:

- La información de apoyo para los antecedentes;
- Resúmenes de las características de los estudios (tales como descripciones detalladas de intervenciones o resultados);

Ver también
- Las tablas adicionales se discuten en el Capítulo 11 (Sección 11.6).

4.7 Estudios y referencias

4.7.1 Referencias a los estudios

Los estudios están organizados bajo cuatro títulos fijos:

Estudios incluidos
Son los estudios que cumplen con los criterios de elegibilidad y están incluidos en la revisión.

Estudios excluidos
Son los estudios que no cumplen con los criterios de elegibilidad y están excluidos de la revisión.

Estudios pendientes de clasificación
Son estudios relevantes que han sido identificados pero que no pueden ser evaluados para la inclusión hasta que se obtengan datos o información adicionales.

Estudios en marcha
Son estudios que están en elaboración y que cumplen (o parecen cumplir) con los criterios de elegibilidad.

Cada uno de estos capítulos puede incluir múltiples estudios (o ningún estudio). Un estudio se identifica por 'ID del estudio' (que generalmente comprende el apellido del primer autor y el año de la principal referencia del estudio). Un año puede estar explícitamente asociado con cada estudio (generalmente el año en que fue terminado, o el año de publicación de la principal referencia) y pueden ser identificadores como un Número Estándar Internacional de Ensayo Controlado Aleatorizado (ISRCTN, sigla en inglés). Además, a cada estudio se le deberá asignar una de las siguientes categorías de 'Fuente de datos':

- Solamente datos publicados;
- Datos publicados y no publicados;
- Solamente datos no publicados;
- Solamente datos publicados (no publicados pedidos pero no utilizados).

Cada estudio puede tener múltiples referencias. A cada referencia se le pueden asignar identificadores tales como un ID de MEDLINE o un DOI. Una referencia para cada estudio debería tener el nivel de 'Referencia principal'. Por seguridad, los autores deberán revisar todas las referencias.

4.7.2 Otras referencias

Las referencias bibliográficas diferentes de los estudios se dividen en las siguientes dos categorías. Hay que hacer notar que RevMan también incluye una categoría de "Pendientes de clasificación" para facilitar la organización de las referencias mientras se prepara una revisión. Todas las referencias deberían moverse a esta categoría antes de que una revisión quede marcada para enviarla al CDSR, dado que las referencias clasificadas en esta categoría no se publicarán.

Referencias adicionales

Otras referencias citadas en el texto deberán enumerarse aquí, incluidas las citadas en las secciones de Antecedentes y Métodos. Si un informe de un estudio se cita en el texto por otra razón diferente de la referencia al estudio (por ejemplo, por alguna información de

antecedentes o metodológica en la referencia) debe ser mencionada aquí, lo mismo que en los estudios relevantes.

Otras versiones publicadas de esta revisión

Las referencias a otras versiones publicadas de la revisión en una revista, un libro de texto, CDSR o en cualquier otro sitio deberían ser listadas aquí. Como regla general, solo se deberían listar las nuevas versiones de citación de las publicaciones en el CDSR (por ej. aquellas que han generado un nuevo registro en MEDLINE).

Por seguridad, los autores deberían revisar todas las referencias.

4.8 Datos y análisis

Los resultados de los estudios incluidos en una revisión están organizados en un orden: los estudios se acogen dentro de subgrupos (opcional), que a su vez se integran dentro de los resultados, y que finalmente son incluidos dentro de las comparaciones (Ver figura 4.8.a). Un estudio puede ser incluido varias veces dentro de los análisis.

El RevMan automáticamente genera gráficos de bosque (*forest plots*) ilustrativos, los cálculos del efecto y los resultados de metanálisis (cuando se seleccionan) a partir de los datos introducidos en la estructura 'Datos y análisis'. El autor puede controlar si se han realizado los metanálisis y cómo se realizaron.

Nota: Los 'Datos y análisis' deberían considerarse información suplementaria porque puede que no aparezcan en algunos formatos de la revisión publicada. Pueden seleccionarse *forest plot* clave (que contienen datos para cada estudio) para incluirlos siempre con el texto completo de la revisión al seleccionarlos como figuras (ver Sección 4.9). Sin embargo, la revisión Cochrane publicada completa en la *CDSR* contendrá toda la sección de 'Datos y análisis' como series de *forest plot* o tablas.

Los autores deberán evitar incluir comparaciones o resultados de los que no haya datos (es decir, tener *forest plot* sin estudios). En cambio, los autores deberán anotar en el texto de la revisión que no hay datos disponibles para las comparaciones. Sin embargo, si la revisión tiene una tabla de 'Resumen de hallazgos', los principales resultados deben incluirse en ésta, sin tener en cuenta si los datos están disponibles en los estudios incluidos.

Comparación

Las comparaciones deberán corresponder a las preguntas o hipótesis de los 'Objetivos'.

Desenlaces

Son posibles cinco tipos de desenlaces: datos dicotómicos, datos continuos, estadísticos 'O – E' y 'V', varianza inversa genérica (estimación puntual y error estándar) y otros datos (sólo texto).

Subgrupo

Los subgrupos pueden relacionarse con subconjuntos de estudios (por ejemplo, ensayos que utilizan diferentes duraciones de fisioterapia) o con una subdivisión del desenlace (por ejemplo, corto plazo, medio plazo y largo plazo).

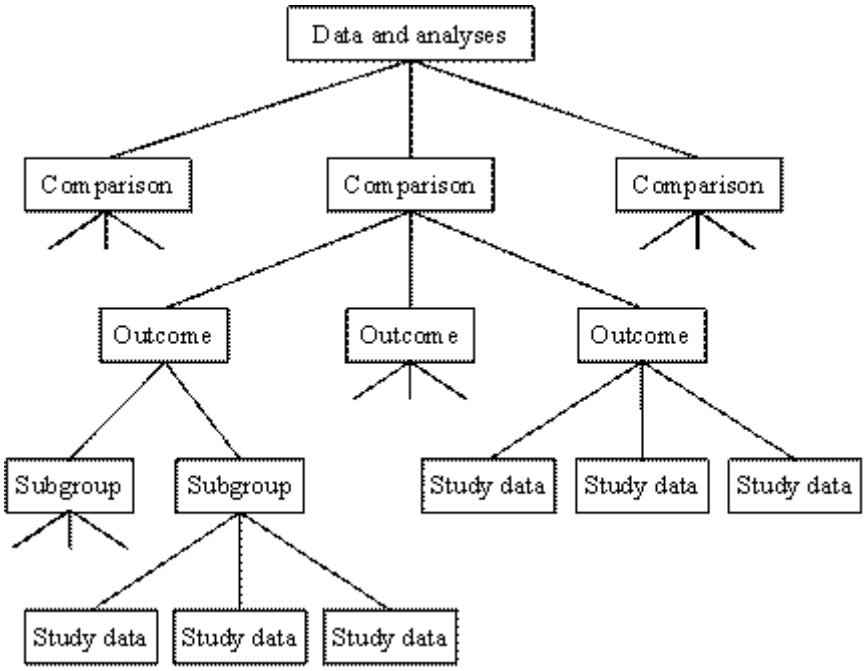
Datos del estudio

Los datos para cada estudio deben introducirse en un formato particular, específico para el tipo de dato de desenlace (por ejemplo, tamaño de la muestra, desviación estándar y promedio para cada grupo de datos continuos).

Ver también

- Los análisis están clasificados en el Capítulo 9: incluyen discusión de comparaciones (Sección 9.1.6), tipos de datos de resultado (Sección 9.2) y subgrupos (Sección 9.6). Las conversiones útiles de los datos informados al formato requerido se encuentran en el Capítulo 7 (Sección 7.7).

Figura 4.8.a: Ilustración de la jerarquía de la sección 'Datos y análisis'.



4.9 Figuras

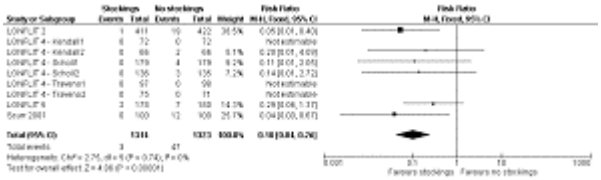
Dentro del texto de la revisión pueden incluirse cinco tipos de figuras (Ver Tabla 4.9.a). Estas figuras se presentaran siempre con el texto completo de la revisión. Cada figura debe tener un título que aporte una descripción breve (o explicación) de la figura, y debe relacionarse (mediante un enlace) con el texto de la revisión.

Ver también

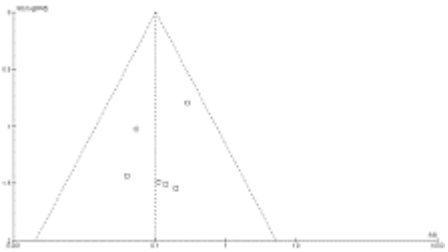
- Los temas de la selección de figuras se discuten en el Capítulo 11 (Sección 11.4.2).

Tabla 4.9.a: Tipos de figuras que pueden incluirse en una revisión Cochrane

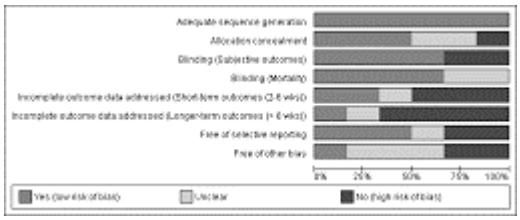
RevMan forest plot



RevMan funnel plot

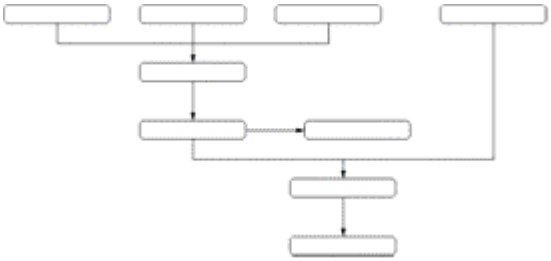


RevMan ‘risk of bias’ graph

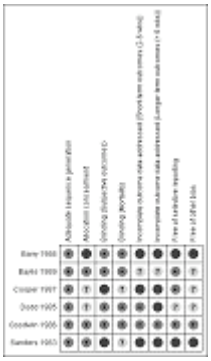


RevMan study flow diagram (PRISMA template)

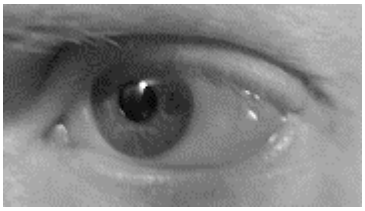
RevMan flow diagram



RevMan ‘risk of bias’ summary



Other figure



4.9.1 Gráficos del RevMan

Se pueden seleccionar los gráficos de árbol (*forest plots*) y los de tunel (*funnel plots*) de los disponibles en "Datos y análisis". Las representaciones gráficas de las valoraciones acerca del riesgo de sesgo pueden generarse también con el RevMan y quedar incluidas como figuras. Finalmente los diagramas de flujo se pueden generar también con el RevMan (versión 5.1 en adelante) para describir el flujo de registros y estudios de la revisión. Hay dos tipos de estos diagramas de flujo: el primero sigue exactamente el formato de PRISMA y su estructura no se puede alterar; el segundo es una herramienta flexible para crear una estructura específica del diagrama de flujo.

Ver también

- Los gráficos de árbol se discuten en el Capítulo 11 (Sección 11.3.2).
- Los gráficos de tunel se discuten en el Capítulo 10 (Sección 10.4).
- Los gráficos de "Riesgo de sesgo" y los resúmenes de "Riesgo de sesgo" se discuten en el Capítulo 8 (Sección 8.6).
- Los diagramas de flujo se discuten en el Capítulo 11 (Sección 11.2.1).

4.9.2 Otras figuras

Los gráficos y otras imágenes que no son generadas por el RevMan pueden incluirse como figuras. Éstas nunca deben usarse como contenido que pueda ser generado de otras maneras dentro del RevMan, por ejemplo como *forest plot* o como tablas adicionales.

Los autores son responsables de obtener el permiso para incluir imágenes en la revisión y asegurarse de que las imágenes coinciden con la publicación. Si se consigue el permiso para publicar una imagen que tiene registro de propiedad, la frase final del pie de la figura debe ser: "Derechos de propiedad © [año] [nombre del propietario del registro, u otro texto apropiado]: reproducido con permiso".

Ver también

- Las figuras que muestran análisis estadísticos deberán seguir la guía correspondiente preparada por el Grupo de Métodos Estadísticos (ver Material adicional en el sitio de la red: www.cochrane.org/resources/handbook).

4.10 Fuentes de financiación de la revisión

Los autores deberán agradecer las ayudas económicas que recibieron para la revisión, y otras formas de ayuda, tales como los apoyos de su universidad o institución por medio de un salario. Las fuentes de apoyo se dividen en 'internas' (aportada por las instituciones en las que se lleva a cabo la revisión), y 'externas' (aportadas por otras instituciones o agencias financieras). Debe mencionarse cada fuente de apoyo, su país de origen, y la clase de apoyo recibido.

4.11 Comentarios

Cada comentario que se incorpora a una revisión publicada se identifica con un título corto y la fecha. El Resumen, la Respuesta y los Colaboradores serán subtítulos de esta sección. El resumen debiera ser preparado por el editor de comentarios del CRG correspondiente, si es necesario, juntamente con la persona que envía el comentario. El (los) autor(es) de la revisión debe(n) preparar la respuesta. Los nombres de las personas que contribuyeron a responder el comentario deben mencionarse en el apartado 'Colaboradores'

Ver también

- Más información sobre los Comentarios se encuentra en el Capítulo 3 (Sección 3.6).

4.12 Anexos

Los anexos constituyen un lugar para la información adicional como:

- Estrategias detalladas de búsqueda (los anexos son los sitios indicados para colocarlas);
- Explicación detallada de métodos estadísticos no estándar;
- Formularios de extracción de datos;
- Detalles de los desenlaces (por ejemplo, escalas de medida).

Los anexos pueden no aparecer en algunos formatos de la revisión publicada.

4.13 Información del capítulo

Editores: Julian PT Higgins y Sally Green.

La versión en inglés de este capítulo debe citarse como: Higgins JPT, Green S (editors). Chapter 4: Guide to the contents of a Cochrane protocol and review. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: Este capítulo se elaboró sobre versiones previas del Manual. Para detalles de los autores y editores anteriores del manual, por favor consulten el Capítulo 1 Sección 1.4). La lista de los títulos recomendados fue desarrollada por Julian Higgins en discusión con Mike Clarke, Sally Hopewell, Jacqueline Birks, varios Coordinadores de Grupos de Revisión, un grupo de evaluación del riesgo de sesgo, y miembros del Grupo Asesor del Manual. Entre los autores que contribuyeron a las actualizaciones recientes se incluyen Ginny Brunton, Mike Clarke, Mark Davies, Frances Fairman, Sally Green, Julian Higgins, Nicki Jackson, Harriet MacLehose, Sandy Oliver, Peter Tugwell y Janet Wale. Agradecemos a Lisa Askie, Sonja Henderson, Monica Kjeldstrøm, Carol Lefebvre, Philippa Middleton, Rasmus Moustgaard y Rebecca Smyth por sus útiles comentarios.

4.14 Referencias

Docherty 1999
Docherty M, Smith R. The case for structuring the discussion of scientific papers. BMJ. 1999;318:1224-1225.

Flanagin 1998
Flanagin A, Carey LA, Fontanarosa PB, Phillips SG, Pace BP, Lundberg GD, Rennie D. Prevalence of articles with honorary authors and ghost authors in peer-reviewed medical journals. JAMA. 1998;280:222-224.

International Committee of Medical Journal Editors 2006
International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication [Updated February 2006]. Available from: http://www.icmje.org (accessed 1 January 2008).

Rennie 1998
Rennie D, Yank V. If authors became contributors, everyone would gain, especially the reader. American Journal of Public Health. 1998;88:828-830.

Rennie 1997
Rennie D, Yank V, Emanuel L. When authorship fails. A proposal to make contributors accountable. JAMA. 1997;278:579-585.

Yank 1999
Yank V, Rennie D. Disclosure of researcher contributions: a study of original research articles in The Lancet. Annals of Internal Medicine. 1999;130:661-670.

PARTE 2:
MÉTODOS GENERALES PARA LAS REVISIONES COCHRANE

ÍNDICE

Capítulo 5: Formular la pregunta de la revisión y desarrollar los criterios para incluir los estudios

Capítulo 6: La búsqueda de estudios

Capítulo 7: Selección de los estudios y obtención de los datos

Capítulo 8: Evaluación del riesgo de sesgo en los estudios incluidos

Capítulo 9: Análisis de los datos y realización de los metanálisis

Capítulo 10: Análisis del sesgo de informe

Capítulo 11: Presentación de los resultados y las tablas ‘resumen de los resultados’

Capítulo 12: Interpretación de los resultados y obtención de las conclusiones

ÍNDICE
CAPÍTULO 5:
FORMULAR LA PREGUNTA DE LA REVISIÓN Y DESARROLLAR LOS CRITERIOS PARA INCLUIR LOS ESTUDIOS

PUNTOS CLAVE

5.1 PREGUNTAS Y CRITERIOS DE ELEGIBILIDAD

5.1.1 Razones para formular preguntas bien estructuradas
5.1.2 Los criterios de elegibilidad

5.2 DEFINIR EL TIPO DE PARTICIPANTES: ¿QUÉ PERSONAS Y POBLACIONES?

Recuadro 5.2.a: Factores a considerar cuando se desarrollan criterios de 'Tipo de participantes'

5.3 DEFINIR LOS TIPOS DE INTERVENCIONES: ¿QUÉ COMPARACIONES HACER?

Recuadro 5.3.a: Factores a considerar en el desarrollo de los criterios para 'Tipos de intervenciones'

5.4 DEFINIR LOS TIPOS DE DESENLACE: ¿QUÉ CRITERIOS DE VALORACIÓN SON LOS MÁS IMPORTANTES?

5.4.1 Enumerar los desenlaces relevantes
5.4.2 Priorizar los desenlaces: principales, primarios y secundarios. Desenlaces de interés relevantes
Recuadro 5.4.a: Factores a considerar cuando se están desarrollando los criterios para los 'Tipos de desenlace'
5.4.3 Desenlaces adversos
5.4.4 Datos económicos

5.5. DEFINIR EL TIPO DE DISEÑO DE ESTUDIO

5.6 DEFINIR EL ALCANCE DE LA PREGUNTA DE UNA REVISIÓN (AMPLIO FRENTE A RESTRINGIDO)

Recuadro 5.6.a: Algunas ventajas e inconvenientes del enfoque amplio frente al restringido

5.7 CAMBIAR LA PREGUNTA DE LA REVISIÓN

5.8 INFORMACIÓN DEL CAPÍTULO

5.9 REFERENCIAS

CAPÍTULO 5: FORMULAR LA PREGUNTA DE LA REVISIÓN Y DESARROLLAR LOS CRITERIOS PARA INCLUIR LOS ESTUDIOS

Editores: Denise O'Connor, Sally Green y Julian PT Higgins.

Puntos clave

- Una revisión claramente definida y orientada comienza con una pregunta bien delimitada. En las revisiones Cochrane las preguntas son planteadas ampliamente como 'Objetivos' de la revisión, y posteriormente se detallam en la sección de 'Criterios para considerar estudios para esta revisión'
- La pregunta de la revisión debe especificar los tipos de población (los participantes), los tipos de intervenciones (y comparaciones), y los tipos de desenlaces que son de interés. La sigla PICO (iniciales en inglés) ayuda a recordar estos conceptos. Estos componentes de la pregunta, con la especificación adicional del tipo de diseño de estudio que será incluido, forman la base de los criterios de elegibilidad preestablecidos para la revisión.
- Las revisiones Cochrane deben incluir todos los desenlaces que puedan ser relevantes para la toma de decisiones y los pacientes, y no incluir desenlaces triviales. Los desenlaces principales deberían limitarse a un número muy pequeño, e incluir tanto desenlaces adversos como favorables.
- Las revisiones Cochrane pueden dedicarse a cuestiones amplias, o pueden estar definidas más estrechamente. Cada modalidad tiene ventajas y desventajas.

5.1 Preguntas y criterios de elegibilidad

5.1.1 Razones para formular preguntas bien estructuradas

Como en cualquier proyecto de investigación, la primera y más importante decisión al preparar una revisión sistemática es determinar su enfoque. Esto se hace mejor si se delimitan claramente las preguntas que la revisión quiere responder. Las preguntas bien formuladas guiarán muchos aspectos del proceso de la revisión, como determinar los criterios de elegibilidad, la búsqueda de estudios, recoger los datos de los estudios incluidos, y la presentación de los resultados (Jackson 1980, Cooper 1984, Hedges 1994). En las revisiones Cochrane, las preguntas se plantean ampliamente como los 'Objetivos' de la revisión y luego se especifican en detalle en la sección de 'Criterios para considerar estudios para esta revisión'. Además de enfocar la orientación de la revisión, los contenidos de estas secciones son utilizados por los lectores en sus evaluaciones iniciales para ver si la revisión realmente está relacionada con los temas que ellos afrontan.

Un planteamiento de los objetivos de la revisión debe comenzar con un enunciado preciso del objetivo principal, preferiblemente en una sola frase. En lo posible el estilo debería seguir la forma: 'Evaluar los efectos de [*intervención o comparación*] para [*problema de salud*] en [*tipos de personas, enfermedad o problema y lugar, si se ha especificado*]'. Esto podría ser seguido por uno o más objetivos secundarios, por ejemplo relacionados con diferentes grupos de participantes, diferentes comparaciones de intervenciones o diferentes medidas de un desenlace.

La especificación detallada de la pregunta de revisión requiere la consideración de varios componentes clave (Richardson 1995, Counsell 1997). La 'pregunta clínica' debería especificar los tipos de población (los participantes), los tipos de intervenciones (y comparaciones), y los tipos de desenlace que son de interés. La sigla en inglés PICO (Participantes, Intervenciones, Comparaciones y *Outcome* (Desenlaces)) sirve de recordatorio. No es necesario hacer el mismo énfasis para referirse a cada uno de estos componentes cuando se delimita la pregunta. Por ejemplo, una revisión podría concentrarse en las diversas opciones de tratamiento para un estadio particular de cáncer de mama, con una definición clara del grado y la severidad de la enfermedad; o centrarse alternativamente en un medicamento determinado, en cualquier estadio del cáncer de mama, con la definición del tratamiento claramente definida.

5.1.2 Los criterios de elegibilidad

Uno de los aspectos que distinguen una revisión sistemática de una revisión narrativa es la especificación previa en la revisión de los criterios para la inclusión y exclusión de estudios (criterios de elegibilidad).

Los criterios de elegibilidad son una combinación de los aspectos de las preguntas clínicas además de la especificación de los tipos de diseño de estudio que han guiado estas preguntas. Los participantes, las intervenciones y las comparaciones en la pregunta clínica generalmente se convierten inmediatamente en criterios de elegibilidad para la revisión. Los desenlaces generalmente no forman parte de los criterios para la inclusión de los estudios: una revisión Cochrane generalmente buscaría todos los estudios rigurosos (por ejemplo, ensayos clínicos aleatorizados) de una comparación de ciertas intervenciones en una determinada población de participantes, independientemente de los desenlaces medidos o reportados. Sin embargo, algunas revisiones deben restringir legítimamente la elegibilidad a algunos desenlaces específicos. Por ejemplo, la misma intervención puede ser estudiada en las mismas poblaciones con distintos propósitos (por ejemplo, terapia hormonal sustitutiva, o aspirina); o una revisión puede dirigirse específicamente a los efectos adversos de una intervención empleada para diferentes condiciones (ver Capítulo 14, Sección 14.2.3).

En las secciones 5.2 a 5.5 presentamos una perspectiva de los componentes clave de las preguntas y los tipos de diseño de estudio, con ejemplos de aspectos útiles a considerar para cada componente y el desarrollo posterior de criterios de elegibilidad para guiar la inclusión de estudios.

5.2 Definir el tipo de participantes: ¿qué personas y poblaciones?

Los criterios para considerar los tipos de población incluidos en una revisión deberían ser suficientemente amplios para agrupar la probable diversidad de estudios, pero también suficientemente justos para asegurar que se puede obtener una respuesta significativa cuando los estudios sean considerados en conjunto. Con frecuencia es útil considerar en dos etapas los tipos de población de interés. Primero, las enfermedades o condiciones de interés deben definirse utilizando criterios explícitos para establecer su presencia o ausencia. Deberían evitarse los criterios que pudieran forzar la exclusión innecesaria de estudios. Por ejemplo, los criterios diagnósticos que se han desarrollado más recientemente —que pueden ser vistos como el *gold standard* actual para diagnosticar la condición de interés— seguramente no se habrán usado en estudios más antiguos. Del mismo modo, es posible que métodos de diagnóstico costosos o muy nuevos no estén disponibles en muchos países o lugares.

Segundo, se deben definir con claridad la población y el ámbito de interés. Esto incluye decidir si interesa un grupo determinado de población, definido por factores como edad, sexo, raza y nivel educativo, o la presencia de una condición especial como angina o disnea. El interés puede centrarse en un ámbito particular como una comunidad, un hospital, una residencia para la tercera edad, un centro para enfermedades crónicas o una consulta externa. El recuadro 5.2.a muestra algunos factores a considerar cuando se desarrollen los criterios para los 'Tipos de participante'.

Para incluir estudios, por lo general el tipo de participantes que interesan determinan directamente los criterios de elegibilidad relacionados con los participantes. Sin embargo, en ocasiones es un desafío establecer previamente normas para los estudios que sólo se refieren parcialmente a la población de interés. Por ejemplo, si el interés está centrado en niños, sería deseable establecer el límite de los 16 años, pero esto no determina una estrategia para participantes que están entre los 12 y los 18 años. El uso de reglas arbitrarias (tales como "más del 80% de los participantes tenían menos de 16 años") no será práctico si no está disponible la información detallada del estudio. Una expresión como "la mayoría de los participantes tenía menos de 16 años" puede ser suficiente. Aunque existe el riesgo de que los sesgos de los revisores afecten las decisiones posteriores de inclusión, esto debe sopesarse con una estrategia de sentido común en la cual las decisiones de elegibilidad estén de acuerdo con los objetivos de la revisión más que con reglas arbitrarias. Las decisiones difíciles deberán documentarse en la revisión, y los análisis de sensibilidad pueden evaluar el impacto de estas decisiones en los resultados de la revisión (ver Capítulo 9, Sección 9.7).

Cualquier restricción con respecto a las características específicas de la población o el ámbito de interés debería basarse en un sólido razonamiento. Como las revisiones Cochrane son ampliamente reconocidas, se justificaría explicar en la revisión la exclusión de estudios basados en las características de la población. Por ejemplo, centrar una revisión de la efectividad del cribado de mamografías en mujeres entre los 40 y los 50 años, puede justificarse en base a la verosimilitud biológica, a las revisiones sistemáticas previamente publicadas, y a la controversia existente. Por otro lado, centrar una revisión en un subgrupo particular de personas basada en su edad, sexo o raza simplemente por intereses personales debe evitarse cuando no hay justificación biológica o sociológica que sustente esta decisión. Cuando no hay seguridad de que haya diferencias importantes de los efectos entre varios subgrupos de personas, puede ser mejor incluir todo lo relacionado con los subgrupos y luego buscar en los análisis si hay diferencias importantes y verosímiles del efecto (ver Capítulo 9, Sección 9.6). Esto debería planificarse desde el principio, considerándolo como un objetivo secundario y no como resultado de la disponibilidad de datos.

Recuadro 5.2.a: Factores a considerar cuando se desarrollan criterios de 'Tipo de participantes'

- ¿Cómo se ha definido la enfermedad/condición de interés?
- ¿Cuáles son las características más importantes que describen esta población (participantes)?
- ¿Existen factores demográficos relevantes (por ejemplo, edad, sexo, raza)?
- ¿Cuál es el ámbito de interés (por ejemplo, hospital, comunidad, etc.)?
- ¿Cómo se determina el diagnóstico?
- ¿Hay otros tipos de personas que deberían excluirse de la revisión (porque podrían reaccionar a la intervención de una manera diferente)?
- ¿Cómo deberán manejarse los estudios que incluyen sólo un segmento relevante de participantes?

5.3 Definir los tipos de intervenciones: ¿qué comparaciones hacer?

El segundo componente clave de una pregunta bien formulada es especificar las intervenciones de interés y las intervenciones con las cuales se van a comparar (comparaciones). En particular, ¿las intervenciones se van a comparar con una intervención de control inactivo (por ejemplo, placebo, ningún tratamiento, cuidado estándar o un control de lista de espera), o con una intervención de control activo (por ejemplo, una variante diferente de la misma intervención, un medicamento diferente, una clase diferente de tratamiento)?

Cuando se especifican intervenciones con fármacos, se deben considerar factores como su preparación, vía de administración, dosis, duración y frecuencia. Para intervenciones más complejas (como intervenciones educativas o comportamentales), deberán definirse los aspectos comunes o centrales de las intervenciones. En general, es útil considerar exactamente qué se está suministrando, con qué intensidad, con qué frecuencia se suministra, quién lo suministra, y si la gente encargada de suministrar la intervención necesita ser entrenada. Los autores de la revisión también deberían tener en cuenta si alguna variación en la intervención (es decir, basada en dosis/intensidad, modo de suministrarla, frecuencia, duración, etc.) es tan importante que pudiera tener diversos efectos significativos en los participantes y en los desenlaces de interés, y por lo tanto que pueda ser conveniente restringirla.

El recuadro 5.3.a define algunos factores a considerar cuando se desarrollan los criterios para los 'Tipos de intervenciones' (y comparaciones).

Recuadro 5.3.a: Factores a considerar en el desarrollo de los criterios para 'Tipos de intervenciones'

- ¿Cuáles son las intervenciones experimentales y de control (comparaciones) de interés?
- ¿Tiene variaciones la intervención (por ejemplo, dosis/intensidad, forma de suministrarla, personal que la suministra, frecuencia de suministro, duración del suministro, momento de suministrarla)?
- Deben incluirse todas las variaciones (por ejemplo, ¿hay una dosis crítica por debajo de la cual la intervención puede ser clínicamente inapropiada?).
- ¿Cómo se manejarán los ensayos que sólo incluyen parte de la intervención?
- ¿Cómo se manejarán los ensayos que incluyen la intervención de interés combinada con otra intervención (cointervención)?

5.4 Definir los tipos de desenlace: ¿qué criterios de valoración son los más importantes?

5.4.1 Enumerar los desenlaces relevantes

El tercer componente clave de una pregunta bien formulada es la definición de los desenlaces de interés más relevantes. En algunas circunstancias, la medida de determinados desenlaces puede ser un criterio para incluir estudios en la revisión, por ejemplo cuando la intervención de interés está dirigida a prevenir un desenlace en particular. No obstante, el reporte de desenlaces en particular debería determinar la inclusión de estudios en una revisión en contadas ocasiones. En particular, no se deberían excluir estudios por no reportar 'datos de utilidad'.

En general, las revisiones Cochrane deberían incluir todos los desenlaces que puedan ser relevantes para los profesionales sanitarios, los pacientes (y usuarios de servicios sanitarios), el público en general, los administradores y los dirigentes políticos, y no deberían incluir resultados publicados en estudios incluidos si son triviales o no tienen significado para quienes toman decisiones. Los desenlaces que se consideran relevantes, y por lo tanto contemplados en una revisión, no tienen porque haber sido reportados en los estudios individuales necesariamente. Por ejemplo la calidad de vida es un desenlace importante, quizá el desenlace más importante, para las personas que deciden si usar o no la quimioterapia para el cáncer avanzado, aún si los estudios disponibles solamente informan sobre la supervivencia (Capítulo 17). Incluir todos los resultados importantes en una revisión destacará las lagunas de conocimiento en la investigación primaria y anima a los investigadores a corregir estas lagunas en futuros estudios.

Los resultados pueden incluir supervivencia (mortalidad), eventos clínicos (ictus o infarto del miocardio), desenlaces comunicados por los pacientes (por ejemplo, síntomas, calidad de vida), eventos adversos, carga de la enfermedad (por ejemplo exigencias a los cuidadores, frecuencia de las pruebas, restricciones en el estilo de vida) y resultados económicos (por ejemplo, coste y uso de recursos). Es importante que los desenlaces usados para evaluar los efectos adversos, tanto como los desenlaces usados para evaluar los efectos deseables o beneficiosos, estén entre los mencionados en la revisión (ver Capítulo 14). Si se van a considerar combinaciones de desenlaces, debe quedar especificado en el texto de la revisión. Por ejemplo, si un estudio no ofrece datos desagregados entre el ictus no fatal y fatal, ¿deberán estos datos incluirse en un metanálisis si la pregunta se relaciona directamente con la muerte por ictus?

Los autores de la revisión deberían considerar cómo pueden medirse los desenlaces, tanto en el tipo de escala a utilizar como en el momento de la medida. Los desenlaces pueden medirse objetivamente (por ejemplo, la tensión arterial, el número de ictus), o subjetivamente como lo haría un clínico, un paciente, o un cuidador (por ejemplo, escalas de discapacidad). Es importante especificar si las escalas de medida han sido publicadas o validadas. Cuando se define el momento de la medida del desenlace, los autores pueden considerar todos los resultados de las medidas tomadas en el tiempo, o solamente incluir en la revisión momentos seleccionados. Una estrategia es agrupar momentos dentro de intervalos previamente especificados, para representar así los desenlaces a corto, medio y largo plazo, y tomar sólo uno de cada estudio para un desenlace determinado. Es importante dar al momento de la medida del desenlace una importancia considerable, ya que puede influenciar los resultados de la revisión (Gøtzsche 2007).

Debido a que las revisiones Cochrane están siendo incluidas cada vez más en las revisiones globales de revisiones (*Overviews of reviews*) (ver Capítulo 22), la homogeneización de los desenlaces por medio de revisiones dirigidas a temas relacionados entre si deberían facilitar este proceso. Puede ser útil para los autores de la revisión considerar las medidas utilizadas en revisiones relacionadas cuando definen el tipo y el momento de la medida en su propia revisión. Además, varias áreas clínicas están desarrollando conjuntos importantes de medidas de desenlace (criterios de valoración) consensuados para ser utilizados en los ensayos clínicos

aleatorizados, y puede ser útil tenerlos en consideración al definir el detalle de la medida de los desenlaces seleccionados para la revisión.

Se pueden utilizar varias fuentes para desarrollar una lista de desenlaces de interés, que incluyen la experiencia de los autores de la revisión, la información de los usuarios de servicios sanitarios y de los grupos asesores (ver el Capítulo 2), y las pruebas científicas en la literatura (incluida la investigación cualitativa acerca de los desenlaces importantes). Más información sobre el uso de la investigación cualitativa para informar la formulación de preguntas de revisión, incluidos los tipos de medidas del desenlace, se puede encontrar en el Capítulo 20.

Aunque deberían incluirse todos los desenlaces importantes en las revisiones Cochrane, los resultados triviales no deberían considerarse. Los autores necesitan evitar recargar, por ser una fuente de confusión, a los lectores con desenlaces que son de poca o ninguna importancia. Además, las medidas de desenlace indirectas o intermedias, como los resultados de laboratorio o de radiología (por ejemplo, pérdida de densidad mineral ósea, como medida indirecta para fracturas en la terapia hormonal sustitutiva), potencialmente pueden confundir y deben evitarse, o interpretarlos con prudencia, porque no pueden predecir con seguridad desenlaces clínicamente importantes. Los desenlaces indirectos pueden aportar información sobre cómo puede funcionar un tratamiento pero no si actualmente funciona. Muchas intervenciones reducen el riesgo para un desenlace indirecto pero no tienen efecto o tienen efectos dañinos en los desenlaces clínicamente relevantes, y algunas intervenciones no tienen efecto en las medidas indirectas pero mejoran los desenlaces clínicos.

5.4.2 Priorizar los desenlaces: principales, primarios y secundarios

Desenlaces de interés relevantes

Una vez se haya elaborado una lista completa de los desenlaces de interés para la revisión, los autores deberán priorizar los desenlaces y seleccionar los de mayor relevancia para la pregunta de la revisión. Los desenlaces principales son los esenciales para la toma de decisiones, y son aquellos que formarían la base para una tabla de “Resumen de los hallazgos”. Esta tabla aporta la información clave acerca de la cantidad de evidencia para las comparaciones y los desenlaces importantes, la calidad de la evidencia y su magnitud del efecto (ver el Capítulo 11, Sección 11.5). No debería haber más de siete desenlaces principales, que generalmente no deberán incluir desenlaces indirectos o intermedios. No deberían ser elegidos en base a ninguna magnitud de efecto anticipada u observada, o porque probablemente hayan sido mencionados en los estudios que se van a revisar.

Desenlaces principales

Los desenlaces primarios de la revisión deberán ser identificados a partir de los desenlaces principales. Son los desenlaces que se esperaría fueran analizados una vez que la revisión identifique estudios relevantes, y las conclusiones acerca de los efectos de las intervenciones que se revisan se basarán en gran parte en estos desenlaces. En general no debería haber más de tres desenlaces primarios, y deberán incluir al menos un desenlace deseable y uno indeseable (para evaluar los efectos beneficiosos y los adversos, respectivamente).

Desenlaces secundarios

Los desenlaces relevantes no seleccionados como principales deberían ser entonces considerados como secundarios. Además, estos desenlaces podrían incluir un número limitado de desenlaces adicionales que la revisión intenta mencionar. Estos desenlaces pueden ser específicos solamente para algunas comparaciones de la revisión. Por ejemplo, las pruebas de laboratorio y otras medidas indirectas no pueden ser consideradas como desenlaces principales, ya que son menos importantes para informar decisiones que las valoraciones clínicas finales, pero pueden ser útiles en explicar el efecto o determinar la integridad de la intervención (ver el Capítulo 7, Sección 7.3.4).

El recuadro 5.4.a resume los principales factores a considerar cuando se están desarrollando los criterios para los 'Tipos de desenlaces'.

Recuadro 5.4.a: Factores a considerar cuando se están desarrollando los criterios para los 'Tipos de desenlace'

- Los desenlaces relevantes, para su inclusión en la tabla de "Resumen de los hallazgos", son aquellos esenciales para la toma de decisiones, y usualmente se focalizan los desenlaces importantes para el paciente.
- Los desenlaces principales son los dos o tres desenlaces, de entre los relevantes, que debería discutir la revisión si se identifican los estudios suficientes, con el fin de llegar a una conclusión acerca de los efectos (benéficos y adversos) de la(s) intervención(es).
- Los desenlaces secundarios incluyen el resto de los desenlaces relevantes para la revisión (diferentes de los principales), más los desenlaces adicionales útiles para explicar los efectos.
- Asegurar que lo desenlaces cubren los efectos adversos.
- Considerar los desenlaces relevantes para todas las personas que toman decisiones, incluyendo datos económicos.
- Considerar el tipo y el momento de la medida de los desenlaces.

5.4.3 Desenlaces adversos

Es importante que las revisiones Cochrane incluyan información tanto de los desenlaces indeseables como de los deseables de las intervenciones evaluadas. Los autores de la revisión deberían considerar cuidadosamente cómo incluirán los datos sobre los desenlaces indeseables en su revisión, y al menos un desenlace indeseable debería haberse definido como una medida de desenlace principal. La evaluación de los efectos adversos se comenta en detalle en el Capítulo 14.

5.4.4 Datos Económicos

Las personas que toman las decisiones necesitan considerar los aspectos económicos de una intervención, tales como si su aplicación llevará a un uso más eficiente de los recursos. Los datos económicos como el uso de los recursos, los costes o la relación coste-efectividad (o una combinación de ellos), deben por lo tanto incluirse como resultados en una revisión. Es útil desglosar las medidas del uso de los recursos y los costes en partidas específicas o categorías. Es útil considerar una perspectiva internacional en la discusión de los costes. Los temas económicos se comentan en detalle en el Capítulo 15.

5.5 Definir el tipo de diseño de estudio

Ciertos diseños de estudios son más apropiados que otros para responder a determinadas preguntas clínicas. Los autores deberían considerar desde el principio qué diseños de estudio pueden aportarles datos fiables con los cuales orientar los objetivos de su revisión.

Como las revisiones Cochrane orientan sus preguntas a los efectos de la atención sanitaria, se centran principalmente en los ensayos clínicos aleatorizados. La aleatorización es la única manera de prevenir diferencias sistemáticas entre las características basales de los participantes en diferentes grupos de intervención, en términos tanto de variables conocidas como desconocidas (o no medidas) (ver el Capítulo 8). En las intervenciones clínicas, la decisión de quién recibe una intervención, y quién no, está influenciada por varios factores, incluidos los factores pronóstico. La evidencia empírica sugiere que, generalmente, los estudios no aleatorizados producen estimaciones de efecto que indican beneficios más extremos de los efectos de la atención sanitaria que los ensayos clínicos aleatorizados. Sin embargo, la extensión, e incluso la dirección, del sesgo son difíciles de predecir. Estos aspectos se comentan

ampliamente en el Capítulo 13, que aporta orientación de cuándo podría ser apropiado incluir estudios no aleatorizados en una revisión Cochrane.

La restricción de muchas revisiones Cochrane a la inclusión de sólo ensayos aleatorizados está motivada también por una consideración práctica. Los esfuerzos de la Colaboración Cochrane para identificar ensayos aleatorizados no han ido a la par con la identificación de otros tipos de estudios. En consecuencia, los estudios incluidos en una revisión que no sean aleatorizados pueden necesitar esfuerzos adicionales para identificar estudios y mantener la revisión actualizada, y podrían incrementar el riesgo de que el resultado de la revisión pudiera ser influenciado por el sesgo de publicación. Este aspecto, y otros relacionados con los sesgos, que son importantes para tener en cuenta cuando se definen los tipos de estudio (por ejemplo, si restringir la elegibilidad de los estudios en base al lenguaje o al estado de la publicación), se comentan en detalle en el Capítulo 10.

Cuando se definen los criterios de elegibilidad también deberían ser considerados aspectos específicos del diseño y la conducción del estudio, incluso si la revisión está restringida a ensayos aleatorizados. Por ejemplo, deberán tomarse decisiones sobre si son elegibles ensayos aleatorizados por grupos (clusters) (Capítulo 16, Sección 16.3) o ensayos clínicos cruzados (Capítulo 16, Sección 16.4), como también los umbrales para la elegibilidad basada en aspectos tales como el uso de un grupo de comparación con placebo, la evaluación con cegamiento de asignación de los desenlaces, o un período mínimo de seguimiento. Siempre habrá una compensación entre los criterios de diseño de estudios restrictivos (que podrían resultar en la inclusión de estudios con bajo riesgo de sesgo, pero cuyo número es escaso), y criterios de diseño más abiertos (que podría resultar en la inclusión de más estudios, pero pueden tener un mayor riesgo de sesgo). Además, criterios excesivamente amplios podrían resultar en inclusión de evidencia errónea. Si por ejemplo, el interés se centra en si una terapia mejora la supervivencia en pacientes con una enfermedad crónica, sería inapropiado considerar estudios de muy corta duración, excepto para explicitar el hecho de que no pueden responder la pregunta de interés.

5.6 Definir el alcance de la pregunta de una revisión (amplio frente a restringido)

Las preguntas que plantea una revisión pueden tener un alcance amplio o restringido. Por ejemplo, una revisión puede plantear una pregunta amplia sobre si los antiagregantes plaquetarios en general son efectivos para prevenir todos los accidentes trombóticos en los humanos. Por otro lado, una revisión podría plantear si un antiagregante plaquetario determinado, como la aspirina, es efectivo para disminuir los riesgos de un episodio trombótico específico, el ictus, en personas mayores con una historia de ictus previo.

Determinar el alcance de la pregunta de una revisión es una decisión que depende de múltiples factores, incluidas la perspectiva relacionada con la relevancia de la pregunta y su potencial impacto; información de apoyo teórico, biológico y epidemiológico; y los recursos disponibles.

Hay ventajas e inconvenientes tanto para las preguntas amplias como para las reducidas, algunas de las cuales se resumen en la Tabla 5.6.a. La validez de las revisiones con un planteamiento amplio puede ser criticada por ‘mezclar manzanas con naranjas’, especialmente cuando una buena evidencia biológica o sociológica sugiere que varias formulaciones de una intervención se comportan de manera muy diferente, o que varias definiciones de la condición o enfermedad que interesa están asociadas a efectos muy diferentes de la intervención.

En la práctica, una revisión Cochrane puede comenzar con un alcance amplio, y luego dividirse en revisiones más delimitadas a medida que la evidencia se acumula y la revisión original se vuelve poco manejable. Esto puede hacerse por razones prácticas y logísticas, por ejemplo para facilitar la actualización, como también para hacerle más fácil a los lectores mantenerse actualizados con sus hallazgos. Los autores con el apoyo de sus Grupos de Revisión deben decidir si hay casos en los que es apropiado dividir una revisión con un enfoque muy amplio en

una serie de revisiones con alcances menos amplios, y cuáles serían los métodos para realizarlo (ver el Capítulo 3, Sección 3.4.4). Si se debe hacer un cambio importante, tal como dividir una revisión amplia en una serie de revisiones con enfoques menos amplios, se necesitará publicar un nuevo protocolo para cada una de las revisiones que la componen, de manera que documente de forma clara sus criterios de elegibilidad.

El desarrollo de revisiones globales Cochrane (*Cochrane Overviews of reviews*) (Capítulo 22, Sección 22.1.1), en la que se resumen múltiples revisiones Cochrane, puede afectar las decisiones de evaluación de las revisiones. La revisiones globales puede resumir múltiples revisiones Cochrane de diferentes intervenciones para la misma enfermedad o problema, o múltiples revisiones de la misma intervención para diferentes tipos de participantes. Puede ser cada vez más deseable planificar una serie de revisiones con un alcance relativamente estrecho, unidas a una Visión General que resuma sus hallazgos.

Tabla 5.6.a: Algunas ventajas e inconvenientes del enfoque amplio frente al restringido

	Enfoque amplio	Enfoque restringido
Selección de los participantes Por ejemplo, inyección de corticosteroides para la tendinitis del hombro (restringido) o inyección de corticosteroides para cualquier tendinitis (amplio)	*Ventajas*: Resumen completo de la evidencia. Capacidad de evaluar la generalizabilidad de los hallazgos mediante los tipos de participantes. *Inconvenientes*: Puede ser más apropiado preparar una Visión General de las revisiones (ver Capítulo 22). La búsqueda, la recogida de datos, el análisis y la redacción pueden requerir más recursos. Hay riesgo de 'mezclar manzanas con naranjas' (heterogeneidad); la interpretación puede ser difícil.	*Ventajas*: Facilidad de manejo para el grupo de revisión; lectura fácil. *Inconvenientes*: La evidencia puede ser escasa. Los hallazgos pueden no ser generalizables a otros lugares o poblaciones. El alcance podría ser elegido por los autores de la revisión para producir un resultado deseado.
Definición de una intervención Por ejemplo, footing supervisado como tratamiento para la depresión (restringido) o hacer cualquier ejercicio como tratamiento para la depresión (amplio)	*Ventajas*: Amplio resumen de la evidencia. Capacidad para evaluar la generalización de los hallazgos por medio de diferentes implementaciones de la intervención.	*Ventajas*: Facilidad de manejo para el equipo de revisión; fácil de leer.

	Inconvenientes: La búsqueda, la recogida de datos, el análisis y la redacción pueden requerir más recursos. Hay riesgo de 'mezclar manzanas con naranjas' (heterogeneidad); la interpretación puede ser difícil.	*Inconvenientes*: La evidencia puede ser escasa. Los hallazgos no pueden generalizarse a otras formulaciones de la intervención. El alcance podría ser elegido por los autores de la revisión para producir un resultado deseado.
Elección de intervenciones y comparaciones Por ejemplo, alarmas para evitar mojar la cama (restringido) o intervenciones para evitar mojar la cama (amplio)	*Ventajas*: Amplio resumen de la evidencia. *Inconvenientes*: Puede ser poco manejable y más apropiado para presentarlo como una Visión General de revisiones (ver Capítulo 22). La búsqueda, la recogida de datos, el análisis y la redacción pueden necesitar más recursos.	*Ventajas*: Facilidad de manejo para el equipo de revisión. Claridad de los objetivos y lectura fácil. *Inconvenientes*: Pueden tener un valor limitado cuando no estén incluidos en una Visión General.

5.7 Cambiar la pregunta de la revisión

Aunque las preguntas deberían plantearse en el protocolo antes de iniciar la revisión sistemática, estas preguntas no deberían convertirse en un corsé que impida la exploración de temas inesperados (Khan 2001). Las revisiones son análisis de datos existentes que están limitados por poblaciones de estudio, ámbitos, detalles de la intervención, criterios de valoración y diseños de estudio previamente elegidos. Generalmente no es posible formular una pregunta para una revisión que se pueda responder sin conocer algunos de los estudios relacionados con la pregunta, y debe quedar claro que las preguntas que una revisión plantea necesitan modificarse a medida que la evidencia se va acumulando en el transcurso de la revisión.

Aunque es de esperarse una cierta fluidez y un replanteamiento de las preguntas de las revisiones a medida que se logra una mayor comprensión de la evidencia, es importante cuidarse de los sesgos que pueden derivarse de modificar las preguntas. Preguntas dirigidas por los datos identificados *ad hoc* en la literatura pueden generar falsas conclusiones basadas en resultados espúreos. Cualquiera de los cambios al protocolo que resulten de replantear la pregunta de la revisión debería documentarse en la Sección 'Diferencias entre el protocolo y la revisión'. Pueden utilizarse análisis de sensibilidad para evaluar el impacto de los cambios en los hallazgos de la revisión (ver Capítulo 9, Sección 9.7). Al replantear las preguntas es útil hacerse los siguientes interrogantes:

- ¿Cuál es la motivación para el replanteamiento de la pregunta clínica?
- ¿Podría haber sido influenciado el replanteamiento por resultados de cualquiera de los estudios incluidos?

- ¿Son apropiadas las estrategias de búsqueda para la pregunta replanteada (especialmente alguna que ya se haya hecho)?
- ¿Son apropiados los métodos de recolección de datos para la pregunta replanteada?

5.8 Información del capítulo

Editores: Denise O'Connor, Sally Green y Julian PT Higgins.

La versión en inglés de este capítulo debe citarse como: O'Connor D, Green S, Higgins JPT (editors). Chapter 5: Defining the review question and developing criteria for including studies. In: Higgins JPT, Green S (editors), *Cochrane Handbook of Systematic Reviews of Intervention.* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: Esta sección está basada en versiones anteriores del *Manual*. Para detalles sobre los autores y editores anteriores del *Manual*, ver Capítulo 1 (Sección 1.4).

5.9 Referencias

Cooper 1984
Cooper HM. *The problem formulation stage*. In: Cooper HM (editors). Integrating Research: a Guide for Literature Reviews. Newbury Park (CA): Sage Publications, 1984.

Counsell 1997
Counsell C. Formulating questions and locationg primary studies for inclusion in systematic reviews. *Annals of Internal Medicine* 1997; 127: 380-387.

Gøtzsche 2007
Gøtasche PC, Hróbjartsson A, Maric K, Tendal B. Data extraction errors in meta-analyses that use standardized mean differences. *JAMA* 2007; 298: 430-437.

Hedges 1994
Hedges LV. *Statistical considerations*. In: Cooper H, Hedges LV (editors). The Handbook of Research Synthesis, New York (NY): Russell Sage Foundation, 1994.

Jackson 1980
Jackson GB. Methods for integrative reviews. *Review of Educational Research* 1980; 50: 438-460.

Khan 2001
Khan KS, ter Riet G, Clanville J, Sowden AJ, Kleijnen J (editors). *Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for those Carrying Out or Commissioning Reviews* (CRD Report Number 4) (2nd edition). York (UK): NHS Centre for Reviews and Dissemination, University of York, 2001.

Richardson 1995
Richardson 1995 WS, Wilson MS, Nishikawa J, Hayward RSA. The well-built clinical question: a key to evidence based decisions. *ACP Journal Club* 1995: A12-A13.

ÍNDICE

CAPÍTULO 6: LA BÚSQUEDA DE ESTUDIOS

PUNTOS CLAVE

6.1 INTRODUCCIÓN

6.1.1 Aspectos Generales
- 6.1.1.1 Papel de los Coordinadores de la Búsqueda de Ensayos (TSC)
- 6.1.1.2 Minimizar los sesgos
- 6.1.1.3 Estudios frente a publicaciones de estudios
- 6.1.1.4 Derechos de autor y patente

6.1.2 Resumen de conceptos

6.2 FUENTES DE BÚSQUEDA

6.2.1 Bases de datos bibliográficas
- 6.2.1.1 Bases de datos bibliográficas-introducción general
- 6.2.1.2 El Registro Central Cochrane de Ensayos Controlados (CENTRAL)
- 6.2.1.3 MEDLINE y EMBASE
- 6.2.1.4 Bases de datos nacionales y regionales
- Recuadro 6.2.a: Ejemplos de bases de datos bibliográficas electrónicas regionales
- 6.2.1.5 Bases de datos de temas específicos
- Recuadro 6.2.b: Ejemplos de bases de datos bibliográficas electrónicas sobre temas específicos
- 6.2.1.6 Índices de Citas
- 6.2.1.7 Bases de datos de tesinas y tesis
- Recuadro 6.2.c: Ejemplos de bases de datos de tesinas y tesis
- 6.2.1.8 Bases de datos de literatura gris

6.2.2 Revistas y otras fuentes que no son bases de datos bibliográficas
- 6.2.2.1 La búsqueda manual
- 6.2.2.2 Revistas con texto completo disponibles electrónicamente
- Recuadro 6.2.d: Ejemplos de revistas con texto completo disponibles gratis en todo el mundo
- 6.2.2.3 Tablas de contenido
- Recuadro 6.2.e: Ejemplos de organizaciones que ofrecen los servicios de Tabla de Contenidos (TOC)
- 6.2.2.4 Resúmenes y Actas de Conferencias
- Recuadro 6.2.f: Ejemplos de resumen de conferencia especializada como fuentes
- 6.2.2.5 Otras revisiones, guías y listas de referencia como fuentes de estudios
- Recuadro 6.2.g: Ejemplos de guías basadas en la evidencia
- 6.2.2.6 Búsqueda en la web

6.2.3 Estudios no publicados y estudios en proceso
- 6.2.3.1 Registros de Ensayos a nivel Nacional e internacional
- Recuadro 6.2.h: Ejemplos de registros de ensayos a nivel nacional e internacional
- 6.2.3.2 Registros de ensayos de temas específicos
- 6.2.3.3 Registros de ensayos de la industria farmacéutica
- Recuadro 6.2.i: Ejemplos de registros de ensayos de la industria farmacéutica
- 6.2.3.4 Registros de resultados de ensayos y otras fuentes
- Recuadro 6.2.j: Ejemplos de registros de resultados de ensayos

6.2.4 Puntos de resumen

6.3 PLANIFICAR EL PROCESO DE BÚSQUEDA

6.3.1 Incluir a los Coordinadores de la Búsqueda de Ensayos y a los bibliotecarios sanitarios en el proceso de búsqueda
6.3.2 Iniciativas de búsqueda de amplia colaboración
Figura 6.3.a: Ilustración de los contenidos de CENTRAL
6.3.2.1 Qué hay en el Registro Central Cochrane de Ensayos Controlados (CENTRAL) procedente de MEDLINE?
Recuadro 6.3.a: Definiciones y criterios Cochrane para ensayos controlados aleatorizados (RCTs) y ensayos clínicos controlados (CCTs)
Recuadro 6.3.b: Definiciones de la Biblioteca Nacional de Medicina de los Estados Unidos 2008, para los términos de Tipos de Publicación 'Ensayo Controlado Aleatorizado' y 'Ensayo Clínico Controlado'
6.3.2.2 ¿Qué hay en el Registro Central Cochrane de Ensayos Controlados (CENTRAL) procedente de EMBASE?
6.3.2.3 ¿Qué hay en el Registro Central Cochrane de Ensayos Controlados (CENTRAL) procedente de otras bases de datos y de búsqueda manual?
6.3.2.4 ¿Qué hay en el Registro Central Cochrane de Ensayos Controlados (CENTRAL) procedente de los Registros Especializados de los Grupos y Áreas Cochrane de Revisión?
6.3.3 Consultar CENTRAL, MEDLINE y EMBASE: asuntos específicos
6.3.3.1 Consultar el Registro Central Cochrane de Ensayos Controlados (CENTRAL): asuntos específicos
Recuadro 6.3.c: Ejemplo de exclusión de registros de MEDLINE y EMBASE cuando se consulta CENTRAL
6.3.3.2 Consultar MEDLINE y EMBASE: asuntos específicos
6.3.4 Puntos de resumen

6. 4 DISEÑAR ESTRATEGIAS DE BÚSQUEDA

6.4.1 Diseñar estrategias de búsqueda-una introducción
6.4.2 Estructura de una estrategia de búsqueda
6.4.3 Proveedores de servicio e interfases de búsqueda
6.4.4 Sensibilidad versus precisión
6.4.5 Vocabulario controlado y palabras de texto
6.4.6 Sinónimos, términos relacionados, variaciones ortográficas, abreviaturas y comodines
6.4.7 Los operadores Booleanos (AND, OR y NOT)
Figura 6.4.a: Combinar conceptos como grupos de búsqueda
6.4.8 Operadores cercanos (NEAR, NEXT y ADJ)
6.4.9 Lenguaje, fecha y restricciones del formato del documento
6.4.10 Identificar estudios falsos, otras publicaciones retiradas, errores de imprenta y comentarios
6.4.11 Filtros de búsqueda
6.4.11.1 Las estrategias Cochrane de Búsqueda Altamente Sensible para identificar ensayos aleatorizados en MEDLINE
Recuadro 6.4.a: Estrategia Cochrane de Búsqueda Altamente Sensible para identificar ensayos aleatorizados en MEDLINE: versión de máxima sensibilidad (revisada en 2008); formato PubMed
Recuadro 6.4.b: Estrategia Cochrane de Búsqueda Altamente Sensible para identificar ensayos aleatorizados en MEDLINE: versión de máxima sensibilidad y precisión (revisada en 2008); formato PubMed
Recuadro 6.4.c: Estrategia Cochrane de Búsqueda Altamente Sensible para identificar ensayos aleatorizados en MEDLINE: versión de máxima sensibilidad (revisada en 2.008); formato Ovid
Recuadro 6.4.d: Estrategia Cochrane de Búsqueda Altamente Sensible para identificar ensayos aleatorizados en MEDLINE: versión de máxima sensibilidad y precisión (revisada en 2.008); Formato de Ovid
6.4.11.2 Filtros de búsqueda para identificar ensayos aleatorizados en EMBASE
6.4.12 Actualizar las búsquedas

6.4.13 Demostración de las estrategias de búsqueda
Recuadro 6.4.e: Demostración de la estrategia de búsqueda para CENTRAL, para el tema 'Tamoxifeno para cáncer de mama'
Recuadro 6.4.f: Demostración de estrategia de búsqueda para MEDLINE (formato Ovid), en el tema 'Tamoxifeno para cáncer de mama'
6.4.14 Puntos de resumen

6.5 EL MANEJO DE LAS REFERENCIAS

6.5.1 El software bibliográfico
6.5.2 Qué campos descargar
Recuadro 6.5.a: Etiquetas de campo importantes en PubMed
6.5.3 Puntos de resumen

6.6 DOCUMENTAR E INFORMAR EL PROCESO DE BÚSQUEDA

6.6.1 Documentar el proceso de búsqueda
6.6.2 Informar el proceso de búsqueda
6.6.2.1 Informar el proceso de búsqueda en el protocolo
6.6.2.2 Informando el proceso de búsqueda en la revisión
6.6.3 Puntos de resumen

6.7 INFORMACIÓN DEL CAPÍTULO

Recuadro 6.7.a: El Grupo Cochrane de Métodos de Recuperación de Información

6.8 REFERENCIAS

CAPÍTULO 6: LA BÚSQUEDA DE ESTUDIOS

Autores: Carol Lefebvre, Eric Manheimer y Julie Glanville, a nombre del Grupo Cochrane de Métodos para Recuperación de la Información.

Puntos Clave

- Los autores de la revisión deberían trabajar muy de cerca, desde el comienzo, con el Coordinador de Búsqueda de Ensayos (TSC, sigla en inglés) de su Grupo Cochrane de Revisión (CRG, sigla en inglés).
- Las revisiones Cochrane incluyen estudios (no las publicaciones de estos estudios), pero identificar las publicaciones de los estudios es normalmente la forma más conveniente de lograr la identificación de la mayoría de los estudios, y de obtener información acerca d ellos y de sus resultados.
- Los registros públicos de ensayos y los registros de los resultados de los ensayos son una creciente e importante fuente de información.
- El Registro Central Cochrane de Ensayos Controlados (CENTRAL), MEDLINE y EMBASE (si está disponible el acceso, bien para el autor de la revisión o para el TSC) debería consultarse para todas las revisiones Cochrane, bien directamente o a través del Registro Especializado del Grupo Cochrane de Revisión.
- Las búsquedas debieran buscar una sensibilidad alta, lo que puede resultar en una baja precisión.
- Deben evitarse demasiados conceptos *diferentes* (diferentes componentes de la pregunta) de búsqueda, y en cambio deberían combinarse una amplia variedad de términos de búsqueda con OR dentro de *cada* concepto.
- Deben usarse tanto el texto libre como los descriptores (por ejemplo, MeSH y EMTREE).
- Deben usarse estrategias de búsqueda altamente sensibles (filtros) para identificar los ensayos aleatorizados, como las recientemente revisadas Estrategias Cochrane de Búsqueda Altamente Sensibles para identificar ensayos aleatorizados en MEDLINE. No se deben aplicar estos filtros en CENTRAL.

6.1 Introducción

Los Grupos Cochrane de Revisión (CRGs) son los responsables de proveer a los autores de la revisión las referencias de los estudios que son posiblemente relevantes para su revisión. Los CRGs utilizan un Coordinador de Búsqueda de Ensayos (TSC por su sigla en inglés), dedicado a dar este servicio (ver Sección 6.1.1.1). La información de este capítulo está diseñada para ayudar a los autores que quieran realizar búsquedas suplementarias de estudios y para proveer información previa de tal manera que los autores puedan entender mejor el proceso de búsqueda. En todos los casos los autores de la revisión deben contactar con el Coordinador de Búsqueda de Ensayos de su CRG, antes de comenzar la búsqueda, con el fin de obtener el nivel de apoyo que ellos dan.

Este capítulo también podrá serle útil a los Coordinadores de Búsqueda de Ensayos, tanto a los que tengan poca experiencia como a los más experimentados, que quieran consultar este capítulo como una fuente de referencia.

Este capítulo traza algunos aspectos generales de la búsqueda de estudios; describe las principales fuentes de estudios potenciales; y comenta cómo planificar el proceso de búsqueda, el diseño y la realización de las estrategias de búsqueda, manejar las referencias encontradas durante la búsqueda y documentar e informar correctamente sobre el proceso de búsqueda.

Este capítulo se centra en la búsqueda de ensayos aleatorizados. Sin embargo, muchos de los principios de búsqueda mencionados aquí también podrán aplicarse a diseños de estudios tratados en otros capítulos. Para la revisión de algunos temas, por ejemplo las intervenciones complejas, será necesario adoptar otros abordajes e incluir otros estudios además de los ensayos aleatorizados. Se recomienda a los autores de la revisión buscar la orientación específica de sus CRGs y referirse también a los capítulos relacionados de este *Manual*, tales como el Capítulo 13 para estudios no aleatorizados, el Capítulo 14 para efectos adversos, el Capítulo 15 para datos económicos, el Capítulo 17 para resultados reportados por los pacientes, el Capítulo 20 para investigación cualitativa, y el Capítulo 21 para las revisiones en promoción de la salud y salud pública. Los autores de revisiones que buscan estudios para su inclusión en revisiones Cochrane de rendimiento de pruebas diagnósticas, debieran remitirse al *Manual Cochrane de Revisiones Sistemáticas de Rendimiento de Pruebas Diagnósticas*

Las numerosas páginas web enumeradas en este capítulo fueron revisadas en Junio de 2008.

6.1.1 Aspectos Generales

6.1.1.1 Papel de los Coordinadores de la Búsqueda de Ensayos (TSC)

El Coordinador de la Búsqueda de Ensayos para cada CRG es el responsable de ayudar a los autores en la búsqueda de estudios para incluirlos en sus revisiones. La amplitud de la ayuda varía de acuerdo a los recursos disponibles para cada CRG, pero puede incluir alguno o todos de los siguientes: aportar estudios relacionados tomados del Registro Especializado de los CRGs (ver Sección 6.3.2.4 para más información), diseñar las estrategias de búsqueda de las bases de datos bibliográficas más importantes, realizar estas búsquedas en las bases de datos disponibles en el CRG, guardar los resultados de las búsquedas y enviárselas a los autores, aconsejar a los autores sobre cómo realizar búsquedas en otras bases de datos y cómo descargar los resultados en su software de manejo de las referencias (ver Sección 6.5). Contacte su Coordinador de Búsqueda de Ensayos antes de comenzar la búsqueda, para obtener el nivel de ayuda ofrecida.

Si en ese momento un CRG no cuenta con un Coordinador de Búsqueda de Ensayos, debe buscar la ayuda de un bibliotecario en salud o un especialista en informática de su localidad, donde haya alguien con experiencia en dirigir búsquedas de revisiones sistemáticas.

6.1.1.2 Minimizar los sesgos

Las revisiones sistemáticas de las intervenciones requieren una búsqueda amplia, objetiva y reproducible de una gama de fuentes, para identificar tantos estudios relacionados como sea posible (dentro del límite de los recursos). Este es un factor clave para distinguir las revisiones sistemáticas de las revisiones narrativas tradicionales y evitar sesgos y ayuda por lo tanto a obtener cálculos reales de efectos.

Una búsqueda sólo en MEDLINE se considera inadecuada. Una revisión sistemática mostró que sólo entre un 30 a un 80% de todos los ensayos aleatorizados conocidos que habían sido publicados se identificaron usando el MEDLINE (dependiendo del área o tema específicos) (Dickersin 1994). Aún si los registros relacionados están en MEDLINE, es difícil recuperarlos (Golder 2006, Whiting 2008). Ir más allá de MEDLINE es importante, no sólo para asegurar que se han identificado tantos estudios relacionados como ha sido posible, sino también para minimizar los sesgos de selección de aquellos que son encontrados. Depender exclusivamente de una búsqueda en MEDLINE puede recuperar una serie de publicaciones no representativos de todas las publicaciones, que podrían haber sido identificadas mediante una búsqueda amplia de varias fuentes.

Las limitaciones de tiempo y presupuesto le imponen al investigador un equilibrio entre la amplitud de la búsqueda y la eficiencia en el uso del tiempo y los recursos, y la mejor forma de lograr este balance es estar alerta, y tratar de minimizar los sesgos, como los sesgos de publicación y de lenguaje que pueden resultar al restringir la búsqueda de diversas maneras (ver el Capítulo 10, Sección 10.2).

6.1.1.3 Estudios frente a publicaciones de estudios

Para las revisiones sistemáticas los estudios son las unidades básicas de interés y análisis. Sin embargo, un solo estudio puede tener más de una publicación sobre él y cada una de éstas puede contener material útil para la revisión (ver el Capítulo 7, Sección 7.2). Para la mayoría de las fuentes citadas en la Sección 6.2, el proceso de búsqueda recuperará publicaciones individuales de estudios; no obstante hay algunas fuentes basadas en el estudio, como bases de datos de registros de ensayos y de resultados de ensayos (ver Secciones 6.2.3.1 a 6.2.3.4).

6.1.1.4 Derechos de autor y patente

La política de la Colaboración Cochrane es que todos los autores de revisión y otros vinculados con la Colaboración deben adherirse a los acuerdos sobre la legislación de derechos de autor y los términos de la patente de las bases de datos. En relación con la búsqueda de estudios, ésta se refiere en particular a adherirse a los términos y condiciones de uso cuando se hacen búsquedas en las bases de datos y descargan registros, y adherirse a la legislación sobre derechos de autor cuando se obtienen copias de artículos. Los autores de revisión deben buscar ayuda en este tema de sus Coordinadores de Búsqueda de Ensayos o bibliotecarios de salud en su localidad, porque la legislación de derechos de autor varía a través de las jurisdicciones y los acuerdos de patentes a través de las instituciones.

6.1.2 Resumen de conceptos

- Los autores de revisión Cochrane deben buscar la ayuda de los Coordinadores de la Búsqueda de ensayos de su Grupo Cochrane de Revisión (CRG) *antes* de comenzar una búsqueda.
- Si el CRG está en ese momento sin un Coordinador de Búsqueda de Ensayos, debe procurar la guía de un bibliotecario sanitario o de un especialista en información de su localidad, donde sea posible encontrar alguien con experiencia en la búsqueda de revisiones sistemáticas.
- Una búsqueda sólo en MEDLINE no se considera adecuada.
- La política de la Colaboración Cochrane es que todos los autores de revisión, y otros vinculados a la Colaboración, deberían adherirse a los términos de las patentes para las bases de datos y sus condiciones de uso, y a la legislación sobre derechos de autor.

6.2 Fuentes de búsqueda

6.2.1 Bases de datos bibliográficas

6.2.1.1 Bases de datos bibliográficas-introducción general

Las búsquedas en las bases de datos bibliográficas relacionadas con la salud son generalmente la forma más fácil, y la que menos tiempo consume, para identificar un conjunto inicial de publicaciones relacionadas con los estudios. Algunas bases de datos bibliográficas, como MEDLINE y EMBASE, incluyen resúmenes de la mayoría de los informes recientes. Un avance importante de estas bases de datos es que pueden ser consultadas electrónicamente, tanto por palabras en el título o resumen como por la utilización de términos de indexación estandarizados, o vocabulario controlado, asignado a cada registro (ver Sección 6.4.5).

La Colaboración Cochrane ha venido desarrollando una base de datos o registro de publicaciones de ensayos controlados denominada El Registro Central Cochrane de Ensayos Controlados (CENTRAL). Es considerada como la mejor fuente individual de publicaciones de ensayos que podría incluirse en las revisiones Cochrane. Las tres bases de datos bibliográficas, generalmente consideradas como las más importantes fuentes de búsqueda de informes de ensayos-CENTRAL, MEDLINE y EMBASE-, se describirán en más detalle en las secciones siguientes.

Las bases de datos están disponibles mediante pago, suscripción, o por el sistema 'pago por uso'. También pueden estar disponibles gratis en el sitio de uso mediante suscripciones nacionales, licencias amplias en instituciones como universidades u hospitales, a través de organizaciones profesionales como parte de acuerdos de conjunto de sus miembros, o gratis en Internet.

También hay numerosas iniciativas internacionales para tener acceso en línea, a bajo coste o gratis, a las bases de datos (y al texto completo en revistas) en Internet. El Acceso a la Red Sanitaria para la Iniciativa en Investigación (HINARI, sigla en inglés) permite el acceso a una amplia gama de bases de datos incluida *La Biblioteca Cochrane* y a cerca de las 4000 revistas más importantes, de una amplia gama de editores en ciencias biomédicas y ciencias sociales relacionadas con profesionales de la salud, en instituciones sin ánimo de lucro, en cerca de 100 países de bajos recursos.

- www.who.int/hinari/en/

La Red Internacional para la Disponibilidad de Publicaciones Científicas (INASP, sigla en inglés) también permite el acceso a una gama amplia de bases de datos, incluida la *Biblioteca Cochrane* y revistas. Los títulos de las revistas disponibles varían de acuerdo a los países. Para más detalles ver

- www.inasp.info/file/68/about-inasp.html

La Información Electrónica de las Bibliotecas (eIFL, sigla en inglés) es una iniciativa similar, basada en un consorcio de bibliotecas para apoyar patentes asequibles de revistas en 50 países de bajos ingresos y en transición en Europa central, del Este, y del Sur-oeste, la antigua Unión Soviética, África, el Medio Oriente y el sur-oeste de Asia. .

- www.eifl.net/cps/sections/about

Para información más detallada acerca de cómo consultar estas y otras bases de datos, ver las Secciones 6.3.3 y 6.4).

6.2.1.2 El Registro Central Cochrane de Ensayos Controlados (CENTRAL)

El Registro Central Cochrane de Ensayos Controlados (CENTRAL) es la fuente más amplia de publicaciones de ensayos controlados. CENTRAL se publica como parte de la *Biblioteca Cochrane* y es actualizado trimestralmente. En Enero de 2008 (Volumen 1, 2008), CENTRAL contiene cerca de 530.000 citas de informes de ensayos y otros estudios potencialmente elegibles para incluirse en las revisiones Cochrane, de los cuales 310.000 informes de ensayos eran de MEDLINE, 50.000 informes adicionales de ensayos eran de EMBASE y los restantes 170.000 eran de otras fuentes como otras bases de datos y de búsqueda manual.

Muchos de los registros en CENTRAL se han identificado a través de búsquedas sistemáticas en MEDLINE y EMBASE, como se describe en las Secciones 6.3.2.1 y 6.3.2.2. Sin embargo, CENTRAL incluye citas de informes de ensayos controlados que no están indexados en MEDLINE, EMBASE u otras bases de datos bibliográficas; citas publicadas en varios idiomas; y citas que están disponibles sólo en memorias de conferencias u otras fuentes a las cuales es difícil acceder (Dickersin 2002). También incluye la información de registros de ensayos y registros de resultados de ensayos (ver Sección 6.2.3).

CENTRAL es gratis para todos los CRGs a través del acceso a la *Biblioteca Cochrane*. La dirección de la web de *La Biblioteca Cochrane* es: http://www.thecochranelibrary.com. Muchas instituciones y organizaciones sanitarias y académicas dan acceso a sus miembros, y en muchos países hay acceso libre para toda la población (por ejemplo, mediante licencias financiadas por la nación o arreglos con los países de bajos ingresos). La información acerca del acceso a la *Biblioteca Cochrane* para determinados países puede hallarse como 'Acceso a la Cochrane' en la parte superior de la página inicial de la *Biblioteca Cochrane*. .

6.2.1.3 MEDLINE y EMBASE

MEDLINE actualmente contiene cerca de 16 millones de referencias a artículos de revistas, desde 1950 en adelante. En la actualidad, están indexadas en MEDLINE 5.200 revistas en 37 idiomas.

- www.nlm.nih.gov/pubs/factsheets/medline.html

PubMed permite el acceso a una versión libre de MEDLINE que también incluye citas actualizadas no indexadas aún en MEDLINE:

- www.nlm.nih.gov/pubs/factsheets/pubmed.html

Además, PubMed incluye registros de revistas que no están indexadas en MEDLINE y registros considerados 'fuera del ámbito' de revistas que están parcialmente indexadas en MEDLINE. Para mayor información de las diferencias entre MEDLINE y PubMed ver:

- www.nlm.nih.gov/pubs/factsheets/dif_med_pub.html

MEDLINE también está disponible en suscripción por una cantidad, por medio de vendedores de bases de datos en línea, como Ovid. El acceso generalmente es gratis para miembros de las instituciones que pagan las subscripciones (por ejemplo, hospitales y universidades).

La Biblioteca Nacional de Medicina de los Estados Unidos (NLM, sigla en inglés) ha desarrollado el Portal NLM, que le permite a los usuarios acceder a MEDLINE o PubMed junto con otros recursos de la NLM simultáneamente, como la base de datos sobre Proyectos de Investigación de los Servicios de Salud (HSRProj, sigla en inglés), Resúmenes de Congresos y el Apartado TOXLINE para citas en toxicología.

- gateway.nlm.nih.gov/gw/Cmd

EMBASE actualmente contiene cerca de 12 millones de registros desde 1974 en adelante. Actualmente en EMBASE hay 4.800 revistas indexadas en 30 idiomas.

- www.info.embase.com/embase_suite/about/brochures/embase_fs.pdf

EMBASE.com es la versión de EMBASE propia de Elsevier que, además de los 12 millones de registros de EMBASE desde 1974 en adelante, también incluye cerca de 7 millones de registros únicos de MEDLINE desde 1966 hasta la fecha, lo que permite que las dos bases de datos sean consultadas simultáneamente.

- www.info.embase.com/embase_com/about/index.shtml

En 2007, Elsevier lanzó EMBASE Classic que ahora permite el acceso a registros digitalizados de revistas impresas de Excerpta Médica (los índices impresos originales de los cuales fue creado EMBASE) desde 1947 hasta 1973.

- www.info.embaseclassic.com/pdfs/factsheet.pdf

EMBASE está disponible solamente por suscripción. Los autores deberían revisar si sus CRG tienen acceso y si no, ver si está disponible a través de la biblioteca local de su institución.

Para orientarse en cómo acceder a MEDLINE y EMBASE para informes de ensayos, ver las Secciones 6.3.3.2, 6.4.11.1 y 6.4.11.2 respectivamente.

Superposición de bases de datos

De las 4.800 revistas indexadas en EMBASE, 1.800 no están indexadas en MEDLINE. Igualmente, de las 5.200 revistas indexadas en MEDLINE, 1.800 no lo están en EMBASE.

- www.info.embase.com/embase_suite/about/brochures/embase_fs.pdf

El grado actual de superposición de referencia varía mucho de acuerdo al tópico, pero los estudios que comparan las búsquedas en las dos bases de datos generalmente han concluido que una búsqueda amplia requiere que ambas bases sean consultadas (Suárez-Almazor 2000). Aunque las búsquedas en MEDLINE y EMBASE tienden a no identificar los mismos conjuntos de referencias, han encontrado que reaparecen cantidades similares de referencias relacionadas.

6.2.1.4 Bases de datos nacionales y regionales

Además de MEDLINE y EMBASE, que son consideradas generalmente como el paradigma internacional en bases de datos en temas de salud, muchos países y regiones producen bases de datos bibliográficas por medios electrónicos, que se concentran en la literatura producida en esas regiones, y que a menudo incluyen revistas y otras informaciones no indexadas en otras partes. El acceso a muchas de estas bases de datos está disponible gratis en Internet. Otras están disponibles solamente por suscripción o por el sistema 'pague a medida que utilice'. Indexar la complejidad y la consistencia varía, como lo hace la sofisticación de la interfase de búsqueda, pero pueden ser una importante fuente de estudios adicionales de revistas no indexadas en otras bases de datos internacionales, como MEDLINE o EMBASE. Algunos ejemplos se incluyen en el Recuadro 6.2.a

Recuadro 6.2.a: Ejemplos de bases de datos bibliográficas electrónicas regionales

Africa: Índice Médico Africano
indexmedicus.afro.who.int/
Australia: Índice Médico Australoasiático (de pago)
www.nla.gov.au/ami/
China: Base de datos de Literatura Médica China (CBM, sigla en inglés) (en Chino)
www.imicams.ac.cn/cbm/index.asp
Mediterráneo oriental : Índice Médico para la Región Oriental del Mediterráneo
www.emro.who.int/his/vhsl/
Europa: PASCAL (de pago)
international.inist.fr/article21.html
India: IndMED
indmed.nic.in/
Korea: KoreaMed
www.koreamed.org/SearchBasic.php
Latino-América y el Caribe: LILACS
bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&base=LILACS&lang=i&Form=F
Sudeste Asiático: Índice Médico para la Región del Sudeste Asiático. (IMSEAR, sigla en inglés)
library.searo.who.int/modules.php?op=modload&name=websis&file=imsear
Ucrania y la Federación Rusa: Panteleimon
www.panteleimon.org/maine.php3
Pacífico Occidental: Índice Médico de la Región del Pacífico Occidental (WPRIM, sigla en inglés)
wprim.wpro.who.int/SearchBasic.php

6.2.1.5 Bases de datos de temas específicos

La búsqueda en bases de datos para temas específicos, además de CENTRAL, MEDLINE y EMBASE, dependerá del tema de la revisión, del acceso a las bases de datos específicas y a consideraciones presupuestarias. La mayoría de las principales bases de datos para temas específicos están disponibles sólo por suscripción, o en el sistema 'pague cuando utilice'. El acceso a las bases de datos parece por lo tanto limitado a las bases de datos disponibles para el Coordinador de Búsqueda de Ensayos, en la sección editorial del CRG, y las que están disponibles en las instituciones de los autores de la revisión. Una selección de las principales bases de datos para temas específicos que pueden estar disponibles a través de subscripciones institucionales (y por tanto libres 'en el sitio de uso') o están disponibles gratis en Internet, se enumeran en el Recuadro 6.2.b, junto con direcciones de Internet para mayor información. Los detalles de acceso varían de acuerdo a la institución. Los autores de la revisión debieran buscar el consejo de sus bibliotecarios de salud locales para el acceso en su institución.

Además de las bases de datos sobre temas específicos, los recursos de búsqueda incluyen:

- Google escolar (gratis en Internet):
    - scholar.google.com/advanced_scholar_search?hl=en&lr=
- Intute (gratis en internet):
    - www.intute.ac.uk/
- Base de datos para Convertir la Investigación en Práctica (TRIP, sigla en inglés) (recurso sanitario basado en la evidencia) (gratis en Internet):
    - www.tripdatabase.com/

Recuadro 6.2.b: Ejemplos de bases de datos bibliográficas electrónicas sobre temas específicos

Biología y farmacología • Resúmenes de Biología / Avances BIOSIS: biosis.org/ • El Archivo Derwent de Medicamentos: scientific.thomson.com/support/products/drugfile/ • Resúmenes Farmacéuticos Internacionales: scientific.thomson.com/products/ipa/
Promoción de la salud • BiblioMap – EPPI-Base de datos central de investigación en promoción de la salud (gratis en Internet): eppi.ioe.ac.uk/webdatabases/Intro.aspx?ID=7 • Base de datos para Promocionar las Revisiones sobre la eficacia en Salud (DoPHER) (gratis en Internet): eppi.ioe.ac.uk/webdatabases/Intro.aspx?ID=2
Salud Internacional • Salud Global: www.cabi.org/datapage.asp?iDocID=169 • POPLINE (salud reproductiva) (gratis en Internet): db.jhuccp.org/ics-wpd/popweb/
Enfermería y Ciencias de la Salud • Medicina Asociada y Complementaria (AMED, sigla en inglés): www.bl.uk/collections/health/amed.html • Índice de Enfermería Británica (BNI, sigla en inglés): www.bniplus.co.uk/ • Índice acumulativo de Enfermería y Ciencias de la Salud (CINAHL, sigla en inglés): www.cinahl.com/ • EMCare: www.elsevier.com/wps/find/bibliographicdatabasedescription.cws_home/708272/description#description • MANTIS (osteopatía y quiropráctica): www.healthindex.com/ • OTseeker (revisiones sistemáticas y ensayos aleatorizados evaluados en terapia ocupacional) (gratis en Internet): www.otseeker.com/ • Base de Datos de la Evidencia en Fisioterapia (PEDro) (revisiones sistemáticas y ensayos aleatorizados evaluados en fisioterapia) (gratis en Internet): www.pedro.fhs.usyd.edu.au/
Salud y bienestar social y comunitario • Línea para personas mayores (gratis en Internet): www.aarp.org/research/ageline/ • Datos sobre niños: www.childdata.org.uk/ • ComunidadPRUDENTE: www.oxmill.com/communitywise/ • Atención social en línea (gratis en Internet): www.scie-socialcareonline.org.uk/ • Resúmenes de Servicios Sociales: www.csa.com/factsheets/ssa-set-c.php
Ciencias Sociales, Educación, Psicología y Psiquiatría • Índice y Resúmenes de las Ciencias Sociales Aplicadas (ASSIA, sigla en inglés): www.csa.com/factsheets/assia-set-c.php • Registro de Ensayos Sociales, Psicológicos, Educativos y Criminológicos de la Colaboración Campbell

(C2-SPECTR, sigla en inglés) (gratis en Internet):
geb9101.gse.upenn.edu/

- Centro para la Información de Recursos en Educación (ERIC, sigla en inglés) (gratis en Internet)
www.eric.ed.gov/
- PsycINFO:
www.apa.org/psycinfo/
- Política y Práctica Social (investigación de la ciencia social basada en la evidencia):
www.ovid.com/site/catalog/DataBase/1859.pdf
- Resúmenes Sociológicos:
www.csa.com/factsheets/socioabs-set-c.php

6.2.1.6 Índices de Citas

El Índice de Citas Científicas / Índice de Citas Científicas Ampliado es una base de datos que enumera los artículos publicados en aproximadamente 6.000 revistas científicas, técnicas y médicas importantes, y los enlaza con los artículos en los que han sido citados (conocida como búsqueda de referencia citada). Está disponible en línea como SciSearch y en Internet como Red de la Ciencia. Esta Red de la Ciencia también está incorporada a la Red del Conocimiento. Puede buscarse como una base de datos igual que MEDLINE. También puede usarse para identificar estudios para una revisión si se elige un artículo inicial conocido y relacionado, y luego se revisa cada uno de los artículos citados en este artículo inicial, para ver si ellos también están relacionados con la revisión.
Es una manera de buscar, a partir de la publicación de un artículo importante relacionado, para identificar artículos adicionales relacionados publicados desde entonces. Los registros también incluyen las referencias enumeradas en el registro original, los que a su vez son otra posible fuente de informes relacionados con el ensayo. La búsqueda de citas es una ayuda importante tanto a la búsqueda en la base de datos como en la manual (Greenhalgh 2005). La información sobre estos productos está disponible en:

- scientific.thomson.com/products/sci/
- scientific.thomson.com/products/wos/
- isiwebofknowledge.com/

Existe una base de datos similar para las ciencias sociales, conocida como el Índice de Citas de las Ciencias Sociales:

- scientific.thomson.com/products/ssci/

En el 2004, Elsevier lanzó una base de datos de citas y resúmenes – Scopus, que incluye 15.000 revistas (de las cuales cerca de 1.200 son revistas de libre acceso) y 500 son actas de conferencias. Contiene cerca de 33 millones de resúmenes, y resultados de cerca de 400 millones de páginas web científicas:

- info.scopus.com/overview/what/

6.2.1.7 Bases de datos de tesinas y tesis

Las tesinas y las tesis generalmente no se indexan en las bases de datos generales, como MEDLINE o EMBASE, pero hay excepciones como CINAHL que indexa tesinas de enfermería. Para relacionar estudios relevantes publicados en tesinas o tesis es aconsejable consultar fuentes específicas de tesinas: ver Recuadro 6.2.c.

Recuadro 6.2.c: Ejemplos de bases de datos de tesinas y tesis

- Base de datos ProQuest para tesinas y tesis: indexa más de 2 millones de tesis doctorales y de tesinas de masters:
  www.proquest.co.uk/products_pq/descriptions/pqdt.shtml
- Índice de Tesis en Gran Bretaña e Irlanda: incluye cerca de 500.000 tesis:
  www.theses.com/
- DissOnline: indexa 50.000 tesis alemanas:
  www.dissonline.de/

6.2.1.8 Bases de datos de literatura gris

Hay muchas definiciones de literatura gris, pero generalmente se entiende como la literatura que no está formalmente publicada en fuentes como libros o artículos de revista. Los resúmenes de conferencias y otra literatura gris han mostrado ser el origen de aproximadamente el 10% de los estudios referenciados en las revisiones Cochrane (Mallett 2002). En una revisión Cochrane de metodología recientemente actualizada, en todos los cinco estudios revisados se encontró que los ensayos publicados tenían un efecto de tratamiento total más grande que los ensayos de la literatura gris (Hopewell 2007b). Por lo tanto, no identificar los ensayos publicados en las actas de conferencias y otros modelos de literatura gris podrían afectar los resultados de de una revisión sistemática.

Los resúmenes de conferencias son una fuente importante de literatura gris y se incluyen en la Sección 6.2.2.4.

La Asociación Europea para la Utilización de Literatura Gris (EAGLE, sigla en inglés), ha cerrado la base de datos 'Sistema para la Información de Literatura Gris (SIGLE, sigla en inglés), que era una de las bases de datos de literatura gris más utilizada. En Francia, el Instituto para la Ciencia y la Tecnología-(INIST) ha lanzado OpenSIGLE, que da acceso a todos los registros iniciales de SIGLE, datos nuevos aportados por los miembros de EAGLE, e información de Greynet.

- opensigle.inist.fr

La base de datos del Consorcio para el Manejo de la Información Sanitaria (HMIC, sigla en inglés) contiene registros del Departamento de los Servicios de Biblioteca e Información, del Ministerio de Salud (HD, iniciales en ingles) en Inglaterra y el Servicio del Fondo Real de Información y Biblioteca. Este incluye todas las publicaciones de DH, incluidas las circulares y los comunicados de prensa. La Fundación Real es una organización de beneficencia sanitaria independiente que trabaja para desarrollar y mejorar el manejo de los servicios sociales y sanitarios. La base de datos es considerada una fuente buena de literatura gris en temas como salud y el manejo del cuidado en la comunidad, desarrollo organizacional, desigualdades en salud, participación de los usuarios, y raza y salud.

- www.ovid.com/site/catalog/DataBase/99.jsp?top=2&mid=3&bottom=7&subsection=10

El Servicio Nacional de Información Técnica (NTIS, sigla en inglés) permite el acceso a los resultados de la investigación, tanto a la financiada como a la no financiada por el gobierno de los Estados Unidos de América, y puede facilitar el texto completo del informe técnico de la mayoría de los resultados obtenidos. Desde 1964, NTIS es gratis en Internet.

- www.ntis.gov/

PsycEXTRA es una base de datos compañera de PsycINFO en psicología, ciencia del comportamiento y salud. Incluye referencias de boletines, revistas, periódicos, informes técnicos y anuales, informes gubernamentales y folletos de los consumidores. PsycEXTRA se diferencia de PsycINFO en su formato, porque incluye resúmenes y citas más el texto completo de una gran parte de los informes. No se superpone la información con PsycINFO.

- www.apa.org/psycextra/

6.2.2 Revistas y otras fuentes que no son bases de datos bibliográficas

6.2.2.1 La búsqueda manual

La búsqueda manual significa una revisión manual página por página del contenido completo de un volumen de la revista o actas de conferencias para identificar todos los informes de ensayos que pudieran elegirse. En las revistas los informes de los ensayos pueden aparecer en artículos, resúmenes, columnas nuevas, editoriales, cartas u otros textos. La búsqueda manual de revistas de salud y de actas de conferencias puede ser un agregado útil para buscar bases de datos electrónicos, al menos por dos razones: 1) no todos los informes de los ensayos se incluyen en las bases de datos bibliográficas electrónicos; 2) aún cuando estén incluidos, no contienen los términos de búsqueda específicos en los títulos o resúmenes, o no son indexados con términos que les permitan ser fácilmente identificados como ensayos (Dickersin 1994). Cada revista anual o acta de conferencia debe ser revisada completa e idóneamente por un buscador manual bien entrenado, buscando todos los informes de ensayos, indistintamente del tema, para que una vez se termine su revisión manual no se requiera volver a hacerlo de nuevo. Una Revisión Cochrane de Metodología ha encontrado que una combinación de búsqueda manual con una revisión electrónica es necesaria para una identificación completa de los informes relacionados publicados en las revistas, aún para aquellos que están indexados en MEDLINE (Hopewell 2007a). Este fue el caso especial de los artículos publicados antes de 1991 cuando aún no existía el término de indexación para ensayos aleatorizados en MEDLINE y para aquellos artículos que están en partes de revistas (tales como suplementos y resúmenes de conferencias) que no son indexados rutinariamente en bases de datos como MEDLINE.

Para la identificación de todos los ensayos publicados la Colaboración Cochrane ha realizado esfuerzos de búsqueda manual extensiva, especialmente a través de sus CRGs, Campos y Centros Cochrane. El Centro Cochrane de los Estados Unidos supervisa el registro prospectivo de todas las posibles búsquedas manuales y mantiene archivos de actividades de búsqueda manual en la Lista Original (Revistas) y la Lista Original (Actas de Conferencias) (ver apps1.jhsph.edu/cochrane/masterlist.asp). Cerca de 3.000 revistas han sido revisadas, o lo están siendo dentro de la Colaboración. Las Listas Originales permiten que el progreso de la búsqueda sea registrado y monitorizado para cada título, y también previene la duplicación de esfuerzos que puede ocurrir si la misma revista o acta de conferencia sea revisada por más de un grupo o individuo.

Las entidades y los autores Cochrane pueden priorizar la búsqueda manual basados en dónde esperan identificar la mayoría de los informes de ensayos. Esta priorización puede ser informada por la búsqueda en CENTRAL, MEDLINE y EMBASE en un área temática e identificar qué revistas aparecen asociadas con la mayoría de las citas recuperadas. La evidencia inicial sugiere que la mayoría de las revistas con un alto número de informes de ensayos están indexadas en MEDLINE (Dickersin 2002) pero esto puede reflejar el hecho de que los colaboradores Cochrane han concentrado inicialmente sus esfuerzos en buscar estas revistas. Por lo tanto, las revistas no indexadas en MEDLINE o EMBASE deberían ser consideradas para búsqueda manual.

Los autores generalmente no esperan hacer una búsqueda manual para sus revisiones pero deberán comentarlo con su Coordinador de Búsqueda de Ensayos si en su caso particular la búsqueda manual de algunas revistas o actas de conferencias podrían ser útiles. Los autores que deseen hacer búsqueda manual de revistas o actas de conferencias deben consultar a sus Coordinadores de Ensayos Clínicos, quiénes pueden determinar si la revista o actas de conferencias ya han sido consultadas, y si no lo han sido pueden registrar la búsqueda en la Lista Original correspondiente y dar formación en la búsqueda manual. El material de

entrenamiento está disponible en la página web del Centro Cochrane de los Estados Unidos (apps1.jhsph.edu/cochrane/handsearcher_res.htm).

Toda la correspondencia relacionada con la iniciación, el progreso y la situación de la búsqueda en una revista o acta de conferencia debe ser entre el Coordinador de la Búsqueda de Ensayos del GCR y el personal del Centro Cochrane de los Estados Unidos.

6.2.2.2 Revistas con texto completo disponibles electrónicamente

El texto completo de un creciente número de revistas está disponible electrónicamente en la modalidad de suscripción o gratis en Internet. Además de aportar un método conveniente de recuperar el artículo completo de registros ya identificados, las revistas con el texto completo también pueden buscarse electrónicamente dependiendo de la interfaz de búsqueda, de un modo similar a la forma en que se pueden buscar los registros de las bases de datos bibliográficas.

Es importante especificar si el texto completo de una revista ha sido buscado electrónicamente. Algunas revistas omiten secciones de la versión electrónica en la versión impresa, por ejemplo cartas, y algunas incluyen artículos adicionales solamente en el formato electrónico.

La mayoría de las instituciones académicas se suscribe a una amplia gama de revistas electrónicas y estas están por lo tanto disponibles gratis en el punto de utilización para los miembros de esas instituciones. Los autores de una revisión deberían buscar consejo sobre el acceso electrónico a las revistas del servicio de biblioteca en su institución local. Algunas asociaciones profesionales permiten el acceso a una cantidad de revistas como parte de los derechos de sus miembros. En algunos países existen convenios similares para los empleados de los servicios de salud por medio de licencias nacionales. También hay varias iniciativas internacionales para permitir el acceso en línea gratis o a bajo coste de revistas con textos completos (y bases de datos) en Internet, como el Acceso a la Red de Salud para las Iniciativas de Investigación (HINARI, sigla en inglés), la Red Internacional para la Disponibilidad de Publicaciones Científicas (INASP, sigla en inglés) y la Información Electrónica para Bibliotecas (eIFL, sigla en inglés). Para mayor información sobre estas iniciativas ver la Sección 6.2.1.1).

Ejemplos de algunas revistas con texto completo que están disponibles gratis en todo el mundo, sin suscripción, se mencionan en el Recuadro 6.2.d.

Se recomienda que se haga y se archive una copia electrónica o impresa de algún artículo posiblemente relacionado que se haya encontrado en revistas de suscripción electrónica, porque la suscripción a la revista puede que no sea a perpetuidad. La revista puede dejar de publicarse o cambiar de editores y el acceso a artículos previamente disponible puede cesar. Lo mismo puede decirse de revistas disponibles gratis en Internet, porque las circunstancias de disponibilidad de determinadas revistas podrían cambiar.

Recuadro 6.2.d: Ejemplos de revistas con texto completo disponibles gratis en todo el mundo

- Central BioMed:
  www.biomedcentral.com/browse/journals/
- Biblioteca Pública de la Ciencia (PLoS):
  www.plos.org/journals/
- Central PubMed (PMC):
  www.pubmedcentral.nih.gov/

Páginas Web que enumeran las revistas que ofrecen acceso libre al texto completo:

- Revistas Médicas Libres:
  freemedicaljournals.com/
- Editorial de Conexión amplia:
  highwire.stanford.edu/lists/freeart.dtl

6.2.2.3 Tablas de contenido

Muchas revistas, aún las que están disponibles solamente por suscripción, ofrecen gratis los servicios de Tabla de contenidos (TOC, sigla en inglés), generalmente a través de alertas por e-mail o descargas RSS. Además, hay numerosas organizaciones que ofrecen los servicios TOC: ver en el Recuadro 6.2.e

Recuadro 6.2.e: Ejemplos de organizaciones que ofrecen los servicios de Tabla de Contenidos (TOC)

- Biblioteca Británica Continua (gratis):
  direct.bl.uk/bld/Home.do
- Biblioteca Británica Continua Plus (suscripción):
  www.bl.uk/reshelp/atyourdesk/docsupply/productsservices/bldplus/
- Dentro de la Biblioteca Británica (será reemplazada por la Biblioteca Británica Plus) (suscripción):
  www.bl.uk/inside
- Contenidos Actuales Conntect (suscripción):
  scientific.thomson.com/products/ccc/
- Biblioteca Electrónica Científica en línea (SciELO) – Brasil (gratis):
  www.scielo.br/

Zetoc (Z39.50 Tabla de Contenidos) (gratis como se especifica abajo): Zetoc permite el acceso a la Tabla de Contenidos Electrónicos de la Biblioteca Británica. Es gratis para los miembros del Comité de Sistemas de Información Conjunta (JISC, sigla en inglés) que ha patrocinado instituciones de educación superior y avanzada en el Reino Unido y todos los Servicios Nacionales de Salud (NHS, sigla en inglés) en Escocia y el Norte de Irlanda:zetoc.mimas.ac.uk/

6.2.2.4 Resúmenes y Actas de Conferencias

Aunque las actas de las conferencias no están indexadas en MEDLINE ni en numerosas bases de datos importantes, están indexadas en la base de datos BIOSIS (http://www.biosis.org/). Cerca de la mitad de los ensayos que se informan en resúmenes de conferencias nunca logran una publicación completa, y los que eventualmente son publicados completos han mostrado ser sistemáticamente diferentes de aquellos que nunca son publicados al completo (Scherer 2007). Por lo tanto, es importante tratar de identificar los estudios posiblemente relacionados publicados en los resúmenes de las conferencias a través de fuentes como bases de datos especialzadas y por búsqueda manual, o buscando electrónicamente esos resúmenes que están disponibles como material impreso, en CD-ROM o en Internet. Muchas actas de las

conferencias se publican como suplementos de revistas. Las fuentes especializadas de resumen de conferencia se enumeran en el Recuadro 6.2.f.

Muchos resúmenes de conferencias se publican gratis en Internet, tales como los de la Sociedad Americana de Oncología Clínica (ASCO, sigla en inglés):

- www.asco.org/ASCO/Meetings

Recuadro 6.2.f: Ejemplos de resumen de conferencia especializada como fuentes

- Resúmenes de Temas Biológicos /RRM (iniciales en inglés Informes, Revisiones, Congresos):
  scientific.thomson.com/products/barrm/
- Dentro de la Biblioteca Británica (será reemplazada por la Biblioteca Británica Directa Plus):
  www.bl.uk/inside
- Biblioteca Británica Directa Plus:
  www.bl.uk/reshelp/atyourdesk/docsupply/productsservices/bldplus
- Actas ISI:
  scientific.thomson.com/products/proceedings/

6.2.2.5 Otras revisiones, guías y listas de referencia como fuentes de estudios

Algunas de las fuentes de referencia más convenientes y obvias para estudios potencialmente relacionados son las revisiones existentes. Copias de revisiones publicadas previamente a, o relacionadas con, el tema de interés deben obtenerse y revisarse para buscar los estudios incluidos (y los excluidos). Tal como la *Base de Datos Cochrane para Revisiones Sistemáticas (CDSR)*, la *Biblioteca Cochrane* incluye la *Base de Datos de Resúmenes de Revisiones de Efectos* (DARE) y la *Base de Datos de Evaluación de la Tecnología en Salud* (Base de datos HTA, sigla en inglés), producidas ambas por el Centro para Revisiones y Difusión (CRD, sigla en inglés) de la Universidad de York, en el Reino Unido. Ambas bases de datos aportan información en revisiones publicadas sobre los efectos de la atención sanitaria. Además de ser publicadas y actualizadas trimestralmente en la *Biblioteca Cochrane*, versiones más actualizadas de estas bases de datos están disponibles gratis en el sitio web del CRD, donde son actualizadas más frecuentemente. Por ejemplo, para el volumen de la Biblioteca Cochrane publicado en Enero de 2007, los registros DARE y HTA fueron aportados por el personal del CRD en Noviembre de 2006. La publicación de Enero de 2007 de la *Biblioteca Cochrane* fue el volumen corriente hasta Abril de 2007, así que los registros de DARE y HTA en la *Biblioteca Cochrane* suelen estar desactualizados por entre dos a cinco meses.

- www.crd.york.ac.uk/crdweb

CRD acostumbraba producir la Base de Datos de Revisiones en Proceso CRD a la que se podía acceder por medio del Registro Nacional de Investigación del Reino Unido (NRR), pero desde que fue cerrada en Septiembre de 2007, los registros de las revisiones en proceso han sido transferidos a la base de Datos HTA.

Las revisiones y guías pueden aportar también información útil sobre las estrategias de búsqueda utilizadas en su desarrollo: ver Recuadro 6.2.g. Los servicios de búsqueda específica basados en la evidencia como Convirtiendo la Investigación en Práctica (TRIP), pueden utilizarse para identificar revisiones y guías. Para ver el rango de las fuentes de revisiones sistemáticas buscadas en TRIP ver:

- www.tripdatabase.com/Aboutus/Publications/index.html?catid=11
- www.guideline.gov

MEDLINE, EMBASE y otras bases de datos bibliográficas pueden también utilizarse para identificar artículos y guías de revisión. En MEDLINE, la revision de artículos más apropiada debería indexarse bajo el término Tipo de Publicación ‘Metanálisis’, que fue introducido en

1993, o 'Revisión', que fue intrioducido en 1966. Las Guías deben indexarse bajo el término Tipo de Publicación 'Guía Práctica', que fue introducido en 1991. EMBASE tiene también un término tesauro 'Revisión Sistemática', que fue introducido en 2003, y 'Guía Practica' que fue introducido en 2004.

Hay una estrategia de búsqueda o filtro, denominada 'Revisión sistemática', en PubMed bajo el enlace 'Dudas Clínicas':

- www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml

Es muy amplio en su alcance y recupera muchas referencias que no son revisiones sistemáticas. La estrategia es la siguiente: "Esta estrategia se concentra en recuperar citas identificadas como revisiones sistemáticas, metanálisis, revisiones de ensayos clínicos, medicina basada en la evidencia, conferencias de desarrollo por consenso, guías, y citas de artículos de revistas especializadas en estudios de revisión de utilidad para los clínicos".

- www.nlm.nih.gov/bsd/pubmed_subsets/sysreviews_strategy.html

Las estrategias de búsqueda o filtros se han desarrollado para identificar revisiones sistemáticas en MEDLINE (White 2001, Montori 2005) y EMBASE (Wilczynski 2007). Estrategias de búsqueda para identificar revisiones sistemáticas en otras bases de datos y para identificar guías se mencionan en el sitio web del Subgrupo de Recursos de Filtros de Búsqueda de Información InterTASC.

- www.york.ac.uk/inst/crd/intertasc/sr.htm

Tanto como buscar las referencias citadas en las revisiones sistemáticas existentes y en los metanálisis, las listas de referencia de los estudios identificados pueden también buscarse para estudios adicionales (Greenhalgh 2005). Ya que los investigadores pueden citar selectivamente estudios con resultados positivos, las listas de referencia deberán usarse con cuidado como una ayuda a otros métodos de búsqueda (ver el Capítulo 10, Sección 10.2.2.3).

Recuadro 6.2.g: Ejemplos de guías basadas en la evidencia

- Consejo Nacional Australiano de Salud e Investigación Médica: Guías para la Práctica Clínica:
  nhmrc.gov.au/publications/subjects/clinical.htm
- Asociación Médica Canadiense – Infobase: Guías para la Práctica Clínica:
  mdm.ca/cpgsnew/cpgs/index.asp
- Guía Nacional para Intercambio de Información (Estados Unidos):
  www.guideline.gov/
- Biblioteca Nacional de Guías (Reino Unido):
  www.library.nhs.uk/guidelinesFinder/
- Grupo de Guías de Nueva Zelanda:
  www.nzgg.org.nz
- Guías Clínicas NICE (Reino Unido):
  www.nice.org.uk/aboutnice/whatwedo/aboutclinicalguidelines/about_clinical_guidelines.js

6.2.2.6 Búsqueda en la web

Hay escasa evidencia empírica sobre el valor del uso general de recursos de Internet, como Google, para identificar posibles estudios (Eysenbach 2001). Los sitios web de los que aportan fondos para la investigación y de los fabricantes de equipos deben ser fructíferos. Los sitios web de búsqueda de la industria farmacéutica pueden ser útiles, particularmente la de registros de ensayos, mencionada en la Sección 6.2.3.3. Si se realizan consultas por Internet se recomienda a los autores de la revisión que impriman una copia o guarden electrónicamente una copia de los detalles de la información acerca de cualquier estudio posiblemente relacionado que se encuentre en Internet, más allá que simplemente marcar el sitio, en caso de que el registro del ensayo se retire o se altere posteriormente. Es importante guardar un registro de la fecha en que el sitio web fue consultado, para poder citarlo.

6.2.3 Estudios no publicados y estudios en proceso

Algunos estudios terminados nunca son publicados. Se ha documentado una asociación entre los resultados 'significativos' y la publicación por medio de una cantidad de estudios, como se resume en el Capítulo 10 (Sección 10.2). Localizar estudios no publicados, e incluirlos en una revisión sistemática cuando son elegibles y apropiados es importante para disminuir los sesgos. No hay una manera fácil y confiable de obtener información sobre estudios que se han completado pero nunca publicado. Esta situación está mejorando como resultado de numerosas iniciativas:

- El Número de Registro Estándar Internacional del Ensayo Clínico Controlado Aleatorizado, es un esquema lanzado como el primer servicio en línea que asigna números únicos a los ensayos controlados aleatorizados en todas las áreas de la sanidad y en todos los países del mundo, y consecuentemente es EnsayosClínicos.gov (ver Sección 6.2.3.1);
- La creciente aceptación por parte de los investigadores de la importancia de registrar los ensayos al comienzo;
- El apoyo al registro desde el comienzo por los editores de las principales revistas médicas y su negativa a publicar posteriormente informes de ensayos no registrados adecuadamente (De Angelis 2004, De Angelis 2005);
- La Política de Acceso Público de los Institutos Nacionales de Salud (NIH, sigla en inglés), de los Estados Unidos, (ver publicaccess.nih.gov/), que hasta Diciembre de 2007 fue voluntario, ahora exige que "todos los investigadores financiados por el NIH envíen o hayan enviado por su cuenta a PubMed Central de la Biblioteca Nacional de Medicina una versión electrónica de sus trabajos finales revisados por pares y aceptados para publicación, que estén disponibles públicamente no más tarde de 12 meses después de la fecha oficial de su publicación".
    - publicaccess.nih.gov/policy.htm

Los colegas pueden ser una fuente importante de información sobre estudios no publicados, y los canales informales de comunicación pueden ser a veces las únicas maneras de identificar datos no publicados. Cartas formales de solicitud de información también se pueden utilizar para identificar estudios terminados pero no publicados. Una manera de hacerlo es enviar una lista amplia de artículos relacionados, junto con los criterios de inclusión para la revisión, al primer autor de los informes de los artículos incluidos preguntándole si conocen algunos estudios adicionales (publicados o no publicados) que puedan tener relación. También puede ser deseable enviar la misma carta a otros expertos y a compañías farmacéuticas o a otros investigadores con interés en ese área. No olvidar que pedirles a los investigadores información sobre estudios terminados pero no publicados no siempre ha sido fructífero (Hetherington 1989, Horton 1997) , aunque algunos investigadores han informado que esta es una manera importante de recuperar estudios para revisiones sistemáticas (Royle 2003, Greenhalgh 2005). Algunas organizaciones asignan sitios web para proyectos de revisiones sistemáticas enumerando los estudios identificados a la fecha e invitan a enviar información de estudios no enumerados aún. También se ha sugerido que legislaciones tales como los Actos de Libertad de Información en países como el Reino Unido y los Estados Unidos, podrían ser usados para ganar acceso a información sobre ensayos no publicados (Bennett 2003, McLean 2003).

También es importante identificar los estudios en proceso, así cuando más tarde la revisión sea actualizada puedan ser evaluados para una posible inclusión. La información acerca de los posibles estudios relacionados que estén en proceso debería incluirse en la revisión en la tabla 'Características de estudios en proceso' (ver el Capítulo 4, Sección 4.6.5). Conocer la existencia de un estudio en proceso posiblemente relacionado también podría afectar decisiones con respecto a cuándo actualizar una revisión específica. Desafortunadamente no existe un registro único, total, centralizado de ensayos en proceso (Manheimer 2002). Sin embargo, se han hecho esfuerzos por numerosas organizaciones, incluidas organizaciones que representan la industria farmacéutica y por las mismas compañías farmacéuticas, para comenzar a dar un acceso central a ensayos en proceso y en algunos casos a resultados de

ensayos que están terminando, tanto a nivel nacional como internacional. En un esfuerzo para mejorar esta situación, la Organización Mundial de la Salud (WHO, sigla en inglés) lanzó el Portal de Búsqueda Plataforma Internacional de Registro de Ensayos Clínicos, en mayo de 2007 para buscar por medio de un rango de registros de ensayos, similar a la iniciativa lanzada unos años antes por Ensayos Controlados Actuales, con su así llamado *meta*-Registro. Actualmente (como en junio de 2008), el portal de la WHO solamente busca por medio de tres registros primarios (El Registro de Ensayos Clínicos de Australia y Nueva Zelandia, Ensayos Clínicos.gov y el Número de Registro Estándar Internacional de Ensayos Controlados Aleatorizados Actualizado). Pero puede anticiparse que se incluirán otros registros a medida que el proyecto progrese.

6.2.3.1 Registros de Ensayos a nivel Nacional e internacional

La el Recuadro 6.2.h enumera los registros de ensayos a nivel nacional e internacional.

Además, Drugs@FDA proporciona información sobre la mayoría de los medicamentos aprobados en los Estados Unidos desde 1939. Para los que han sido aprobados más recientemente (desde 1998), con frecuencia hay una 'revisión' que contiene los análisis científicos que proporcionaron las bases para la aprobación del nuevo medicamento.

- www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm

Otras agencias nacionales y regionales para la aprobación de medicamentos también pueden ser fuentes útiles de información de ensayos.

Recuadro 6.2.h: Ejemplos de registros de ensayos a nivel nacional e internacional

- La Asociación de la Industria Farmacéutica Británica (ABPI, sigla en inglés) – Base de datos de Ensayos Clínicos de la Industria Farmacéutica:
  www.cmrinteract.com/clintrial/
- El Registro de Ensayos Clínicos de Australia y Nueva Zelanda:
  www.anzctr.org.au/
- Centro de Observación del Servicio de Inclusión de Ensayos Clínicos:
  www.centerwatch.com/
- Registro de Ensayos Clínicos de China:
  www.chictr.org/Default.aspx
- Registro de Ensayos Clínicos.gov:
  clinicaltrials.gov/
- Servicio de Información de Investigación y Desarrollo Comunitario (de la Unión Europea) (ensayos y otra investigación):
  cordis.europa.eu/en/home.html
- Ensayos Controlados Actuales *meta*Registro de Ensayos Controlados (*m*RCT, sigla en inglés) – registros activos:
  www.controlled-trials.com/mrct/
- Ensayos Controlados Actuales *meta*Registro de Ensayos Controlados (*m*RCT, sigla en inglés) – registros archivados:
  www.controlled-trials.com/mrct/archived
- Agencia Europea de Medicinas (EMEA):
  www.emea.europa.eu/index/indexh1.htm
- Registro de Ensayos de Alemania – aún no lanzado. Se llegó a un acuerdo final el 30 de Agosto del 2007 – se incluirá bajo el Portal de Búsqueda Plataforma Internacional de Registro de Ensayos Clínicos, de la WHO – para más detalles, siempre y cuando esté disponible, ver:
  www.who.int/trialsearch
- Registro de Ensayos Clínicos de Hong Kong –HKClinicalTrials.com:
  www.hkclinicaltrials.com/
- Registro de Ensayos Clínicos de la India – Registro de Ensayos Clínicos – India (CTRI, sigla en inglés):

www.ctri.in

- El Portal de Búsqueda de la Plataforma Internacional de Registro de Ensayos Clínicos:
  www.who.int/trialsearch
- Portal de Ensayos Clínicos de la Federación Internacional de Asociaciones y Fabricantes Farmacéuticos (IFPMA, sigla en inglés):
  www.ifpma.org/clinicaltrials.html
- Número Internacional Estándar de Ensayos Controlados Aleatorizados:
  www.controlled-trials.com/isrctn/
- Registro de ensayo en Los Países Bajos (Nederlands Trialregister – en Holandés):
  www.trialregister.nl/trialreg/index.asp
- Registro Nacional de Ensayos Clínicos de Sudafrica:
  www.sanctr.gov.za/
- Base de datos de la Oferta de la Red de Investigación Clínica del Reino Unido:
  portal.nihr.ac.uk/Pages/Portfolio.aspx
- Portal de Ensayos Clínicos del Reino Unido:
  www.controlled-trials.com/ukctr/
- Registro Nacional de Investigación del Reino Unido (NRR, iniciales en inglés) (ensayos y otra investigación–archivado en Septiembre de 2007 – ver Portal de Ensayos Clínicos del Reino Unido):
  portal.nihr.ac.uk/Pages/NRRArchive.aspx
- Registro de Ensayos Clínicos (para el Japón) de la Red de Información Médica del hospital universitario (UMIN, sigla en inglés)- UMIN CTR:
  www.umin.ac.jp/ctr/

6.2.3.2 Registros de ensayos de temas específicos

Hay tantos registros de ensayos sobre temas específicos, especialmente en el campo del cáncer-que son demasiados para mencionarlos. Pueden identificarse por búsqueda en Internet y por búsqueda en algunas de la fuentes mencionadas arriba, como el *meta* Registro Actual de Ensayos Controlados de los Ensayos Controlados (*m*RCT, sigla en inglés).

6.2.3.3 Registros de ensayos de la industria farmacéutica

Algunas compañías farmacéuticas tienen disponible la información sobre sus ensayos clínicos por medio de sus propias páginas web, bien sea en lugar-o además de, la información que facilitan por medio de sus sitios web nacionales o internacionales, tales como los mencionados anteriormente. Algunos ejemplos están en el Recuadro 6.2.i

Recuadro 6.2.i: Ejemplos de registros de ensayos de la industria farmacéutica

- Sitio web de los Ensayos Clínicos de AstraZeneca Clinical:
  www.astrazenecaclinicaltrials.com/
- Registro de los Ensayos Clínicos de Bristol-Myers Squibb:
  ctr.bms.com/ctd/registry.do
- Registro de los Ensayos Clínicos de Eli Lilly y Compañía (también incluye los resultados de los ensayos)
  www.lillytrials.com/
- Registro de los ensayos clínicos de GlaxoSmithKline:
  ctr.gsk.co.uk/medicinelist.asp
- Ensayos Clínicos Novartis.com:
  www.novartisclinicaltrials.com/webapp/etrials/home.do
- Registro de los Protocolos de Ensayos Clínicos de Roche:
  www.roche-trials.com/registry.html
- Listados de Ensayos Clínicos de Wyeth:
  www.wyeth.com/ClinicalTrialListings

6.2.3.4 Registros de resultados de ensayos y otras fuentes

Los registros de los resultados de los ensayos terminados son hechos más recientes, que siguieron a los registros de los ensayos en proceso que simplemente enumeran detalles del ensayo. Tienen un valor particular porque los resultados de los ensayos no siempre se publican, y si se publican no siempre se hace de forma completa. Una ley reciente de los Estados Unidos, conocida como la Sección 801 del Acta de Enmiendas de 2007 de la Administración de Drogas y Alimentos (FDAAA 801, sigla en inglés), promulgada en Septiembre de 2007, pidió aumentar Ensayos Clíncos.gov y agregar una base de datos de resultados de ensayos clínicos.

Ejemplos de registros de resultados de ensayos aparecen en la Recuadro 6.2.j.

Además, Resultados de Ensayos Clínicos es un sitio web que alberga presentaciones de muestras de investigadores clínicos que informan los resultados de los ensayos clínicos: www.clinicaltrialresults.org/

Recuadro 6.2.j: Ejemplos de registros de resultados de ensayos

- Portal de Ensayos Clínicos de la Federación Internacional de Asociaciones y Fabricantes Farmacéuticos (IFPMA, sigla en inglés):
  www.ifpma.org/clinicaltrials.html
- Base de datos de Resultados de Estudios Clínicos PhRMA:
  www.clinicalstudyresults.org/about
- Resultados de Ensayos Clínicos de Bristol-Myers Squibb:
  ctr.bms.com/ctd/results.do
- Registro de Ensayos Clínicos de Eli Lilly y Compañía:
  www.lillytrials.com/
- Base de datos de Resultados de Ensayos Clínicos de Roche:
  www.roche-trials.com/results.html
- Resultados de Ensayos Clínicos de Wyeth:
  www.wyeth.com/ClinicalTrialResults

6.2.4 Puntos de resumen

- Los autores de revisiones Cochrane deberán buscar la ayuda de sus Coordinadores de Búsqueda de Ensayos sobre las fuentes que deben consultar.
- CENTRAL está considerada como la mejor fuente individual de informes de ensayos para inclusión en las revisiones Cochrane.
- Las tres bases de datos generalmente consideradas como las fuentes más importantes para buscar estudios para inclusión en las revisiones Cochrane son CENTRAL, MEDLINE y EMBASE.
- De acuerdo con el tema de la revisión, para la búsqueda se deberán seleccionar bases de datos nacionales, regionales y de temas específicos.
- Los resúmenes de conferencias y otra literatura gris pueden ser una fuente importante de estudios para incluir en las revisiones.
- Listas de referencia de otras revisiones, guías, estudios incluidos (y excluidos) y otros artículos relacionados deberán consultarse para estudios adicionales.
- Se deben hacer esfuerzos para identificar estudios no publicados.
- Deberán identificarse y seguirse los ensayos en proceso para su posible inclusión en revisiones al terminarse.
- Registros de ensayos y registros de resultados de ensayos son una fuente importante de ensayos en proceso y no registrados.

6.3 Planificar el proceso de búsqueda

6.3.1 Incluir a los Coordinadores de la Búsqueda de Ensayos y a los bibliotecarios sanitarios en el proceso de búsqueda

Es responsabilidad de cada CRG apoyar a los autores de la revisión para identificar informes de estudios para incluirlos en sus revisiones, y muchos CRGs cuentan con un Coordinador de Búsqueda de Ensayos para realizar esta labor (ver Sección 6.1.1.1). La mayoría de los CRGs ofrecen apoyo a los autores en la identificación de estudios desde las primeras etapas hasta la escritura final de la revisión, para su publicación en la *CDSR*. Este apoyo podría incluir el diseño de estrategias de búsqueda o asesorar sobre su diseño, búsquedas contínuas, en particular en bases de datos no disponibles para el autor de la revisión en su institución, y facilitándole a los autores listas de referencias de estudios del Registro Especializado del CRG y posiblemente de otras bases de datos. La amplitud de los servicios ofrecidos varía entre diferentes CRGs, de acuerdo a los recursos disponibles. Por lo tanto, los autores de la revisión deben animarse a contactar al Coordinador de la Búsqueda de Estudios de su CRG desde el comienzo, para consejo y apoyo.

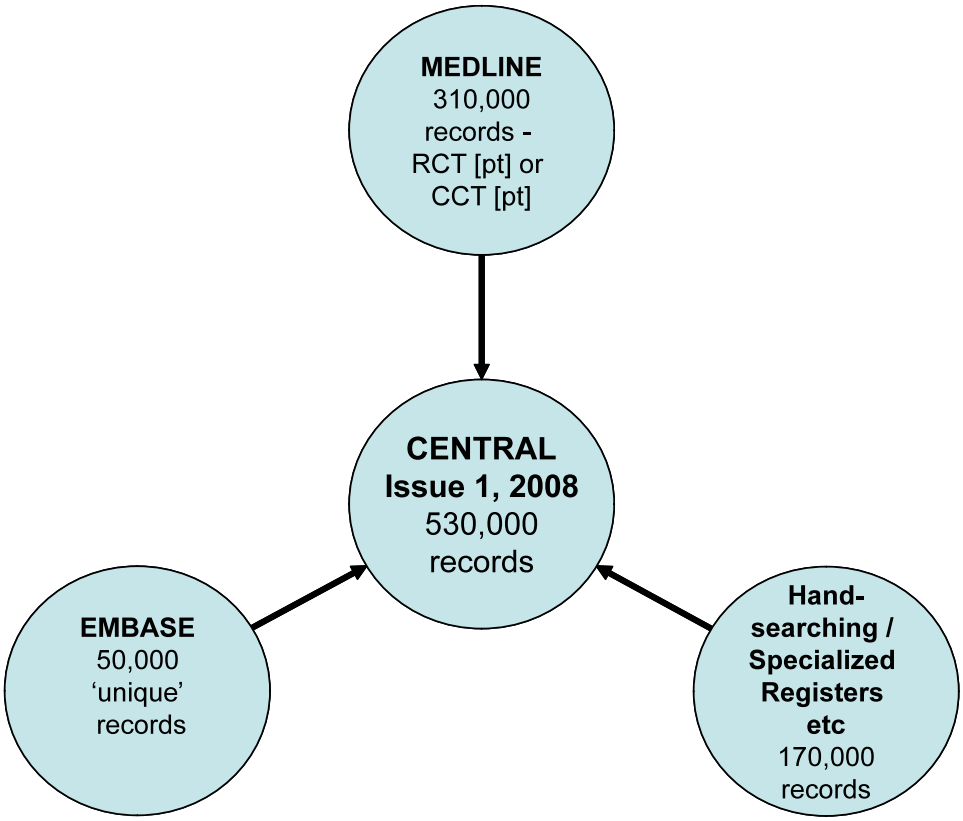
Si los autores dirigen sus propias búsquedas, deberán buscar consejo de su Coordinador de Búsqueda de Ensayos respecto a qué bases de datos consultar y las estrategias precisas a seguir. También deberá tenerse en mente que el proceso de búsqueda necesita documentarse completamente, al menor detalle, para estar seguro de poderlo informar correctamente en la revisión, de tal manera que todas las búsquedas en todas las bases de datos se puedan reproducir. Las estrategias de búsqueda para cada base de datos deberán incluirse completas en un Apéndice. Por lo tanto, los autores de la revisión deberán guardar todas las estrategias de búsqueda y también tomar notas para facilitar completar la sección en el tiempo apropiado. Para más orientación en esto, los autores deberán contactar con su Coordinador de Búsqueda de Ensayos, y ver la Sección 6.6.

Si el CRG normalmente no cuenta con Coordinador de Búsqueda de Ensayos se recomienda que los autores de la revisión busquen la ayuda de un bibliotecario sanitario o especialista en información de preferencia con experiencia en apoyo de revisiones sistemáticas.

6.3.2 Iniciativas de búsqueda de amplia colaboración

Al planificar el proceso de búsqueda es necesario tener en cuenta qué otra búsqueda ya se ha hecho para evitar duplicación innecesaria de esfuerzo. Por ejemplo, durante varios años se han hecho considerables esfuerzos de búsqueda en MEDLINE y EMBASE, y se han incorporado informes de ensayos, desde estas dos grandes bases de datos internacionales, al Registro Central Cochrane de Ensayos Controlados (CENTRAL). Es necesario, por lo tanto, que cualquier búsqueda adicional para una revisión específica tenga en cuenta lo que se ha hecho antes. La Figura 6.3.a ilustra los contenidos de CENTRAL.

Figura 6.3.a: Ilustración de los contenidos de CENTRAL



6.3.2.1 Qué hay en el Registro Central Cochrane de Ensayos Controlados (CENTRAL) procedente de MEDLINE?

CENTRAL contiene todos los registros de MEDLINE indexados con el término Tipo de Publicación 'Ensayo Controlado Aleatorizado' o 'Ensayo Clínico Controlado', que están indexados como estudios en humanos. Estos registros se bajan cada tres meses de MEDLINE por Wiley-Blackwell como parte de CENTRAL para publicación en *La Biblioteca* Cochrane. Para mayor información ver:

- www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/CENTRALHelpFile.html

Un número importante de los registros de MEDLINE codificados cómo 'Ensayo Controlado Aleatorizado' o 'Ensayo Clínico Controlado' en el campo Tipo de Publicación han sido codificados como resultado del trabajo de La Colaboración Cochrane (Dickersin 2002). Los resultados de la búsqueda manual de las entidades Cochrane, de revistas indexadas en MEDLINE se han enviado a la Biblioteca Nacional de Medicina de los Estados Unidos (NLM), donde los registros MEDLINE han sido codificados de nuevo con los tipos de publicación 'Ensayo Controlado Aleatorizado' o 'Ensayo Clínico Controlado', según corresponda. Además, el Centro Cochrane de los Estados Unidos (antes el Centro Cochrane de New England, la Oficina de Providence y el Centro Cochrane de Baltimore) y el Centro Cochrane del Reino Unido han realizado una búsqueda electrónica en MEDLINE desde 1966 hasta 2004, para identificar los informes de ensayos controlados aleatorizados, identificables por los títulos y/o resúmenes, no indexados como tales hasta entonces, utilizando las dos primeras fases de la Estrategia de Búsqueda Cochrane de Alta Sensibilidad, publicada inicialmente en 1994 (Dickersin 1994) y posteriormente actualizada e incluida en el *Manual.* Los términos de texto libre utilizados fueron: ensayo clínico; (sencill$ O dobl$ O agud$ O tripl$) Y (enmascar$ O ceg$); placeb$; aleator$. El signo $ indica el uso de un símbolo truncado. Los siguientes términos de índice de tema utilizados (Título de Tema Médico-MeSH, en inglés) fueron rebatidos: ensayos controlados aleatorizados, asignación aleatoria; método doble ciego;

método simple ciego; ensayos clínicos; placebos. El siguiente título de tema (MeSH) fue utilizado sin discusión: diseño de investigación. Los términos del Tipo de Publicación utilizados fueron: ensayo controlado aleatorizado; ensayo clínico controlado; ensayo clínico.

Se realizó una prueba utilizando los términos en fase tres de la Estrategia de Búsqueda Cochrane de Alta Sensibilidad de 1994, pero la precisión de esos términos, habiendo buscado ya todos los términos en fases una y dos mencionados antes, se consideraron demasiado bajos para garantizar su uso para el proyecto anterior (Lefebvre 2001). Sin embargo, se reconoció que algunos de estos términos podrían ser útiles cuando se combinaran con términos de temas para identificar estudios para algunas revisiones específicas (Eisinga 2007).

La búsqueda anterior se limitó a seres humanos. Los años siguientes fueron completados por el Centro Cochrane de los Estados Unidos (1966-1984; 1998-2004) y por el Centro Cochrane del Reino Unido (1985-1997). Los resultados se han enviado a la NLM y codificados de nuevo en MEDLINE y están incluidos así en CENTRAL. Este proyecto está actualmente en espera. Si el Centro Cochrane de los Estados Unidos puede conseguir fondos para este proyecto podrán continuar la búsqueda electrónica de registros ingresados a MEDLINE en el 2005 y después. Cualquiera de las actualizaciones de esta situación se describirá en el fichero Detalles de Creación CENTRAL en la *Biblioteca Cochrane*:

- www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/CENTRALHelpFile.html

CENTRAL incluye de MEDLINE no sólo informes de ensayos que cumplen la definición más estricta de la Cochrane para un ensayo clínico controlado (Recuadro 6.3.a) sino también informes de ensayos que cumplen las definiciones originales menos estrictas de la NLM (Recuadro 6.3.b), que acostumbraba incluir comparaciones históricas. Actualmente no hay un método que permita distinguir, en CENTRAL o en MEDLINE, cuál de estos registros cumple la definición más estricta de la Cochrane, ya que todos están indexados con el término Tipo de Publicación 'Ensayo Clínico Controlado'.

Recuadro 6.3.a: Definiciones y criterios Cochrane para ensayos controlados aleatorizados (RCTs) y ensayos clínicos controlados (CCTs)

Los registros identificados para inclusión deberán cumplir los criterios de elegibilidad creados y acordados en Noviembre de 1992, publicados inicialmente en 1994 en la primera versión del Manual (ver Capítulo 1, Sección 1.4). De acuerdo con estos criterios de elegibilidad :
Un ensayo es elegible si, con base a la mayor información disponible (generalmente de uno o más informes publicados), se considera que:
• Los individuos (u otras unidades) estudiados en el ensayo fueron definitiva o posiblemente asignados prospectivamente a uno de dos (o más) formas alternativas de atención sanitaria utilizando
Asignación aleatoria o
Algún método cuasi aleatorio de asignación (como alternancia, fecha de nacimiento, o número de registro de caso).

Los ensayos elegibles para inclusión se clasifican de acuerdo al grado de certeza del lector de que se utilizó la asignación aleatoria para formar los grupos comparativos en el ensayo. Si el (los) autor(es) explícitamente declaran (generalmente utilizando alguna variante del término 'aleatorio' apara describir el procedimiento de asignación empleado) que los grupos comparados en el ensayo fueron establecidos por asignación aleatoria, entonces el ensayo se clasifica como RCT (ensayo controlado aleatorizado). Si el (los) autor(es) no declaran explícitamente que el ensayo fue aleatorizado, pero la aleatorización no puede descartarse, el informe se clasificará como un CCT (ensayo

clínico controlado). La clasificación CCT también se aplica a los estudios cuasi-aleatorizados, donde el método de asignación es conocido pero no se considera estrictamente aleatorio, y a ensayos posiblemente cuasi-aleatorizados. Ejemplos de métodos cuasi-aleatorios de asignación incluyen la alternancia, la fecha de nacimiento y el número de registro médico.

La clasificación como RCT o CCT se basa exclusivamente en lo que el autor ha escrito, no en la interpretación del lector; así, no se entiende reflejar una evaluación de la verdadera naturaleza o calidad del procedimiento de asignación. Por ejemplo, aunque los ensayos 'doble ciego' casi siempre son alatorizados, muchos informes de ensayos fallan al no mencionar explícitamente la asignación aleatoria y por lo tanto deben clasificarse como CCT.
Los informes relevantes son los informes de estudios, publicados en cualquier año, que comparan al menos dos formas de cuidado sanitario (tratamientos sanitarios, educación el cuidado de la salud, pruebas o técnicas diagnósticas, una intervención preventiva, etc.) donde el estudio se hace bien en seres humanos o partes de su cuerpo, o partes humanas que serán reemplazadas en seres humanos (por ejemplo, donantes de riñón). Los estudios en cadáveres, dientes extraídos, líneas celulares, etc., no son relevantes. *Los investigadores deberán identificar todos los ensayos clínicos controlados que cumplan estos criterios a pesar de la relevancia para la entidad a la cual están vinculados.*
La proporción más alta posible de todos los informes de ensayos controlados en la atención sanitaria deberá incluirse en CENTRAL. Así, los que buscan en la literatura para identificar ensayos deberán dar a los informes el beneficio de cualquiera de las dudas. Los autores de la revisión decidirán si incluyen un informe determinado en una revisión.

Recuadro 6.3.b: Definiciones de la Biblioteca Nacional de Medicina de los Estados Unidos 2008, para los términos de Tipos de Publicación 'Ensayo Controlado Aleatorizado' y 'Ensayo Clínico Controlado'

Ensayo Controlado Aleatorizado
El trabajo consiste en un ensayo clínico que implica al menos un tratamiento de prueba y un tratamiento de control, registro simultáneo y seguimiento de esos grupos de prueba y de control, y en el que los tratamientos suministrados se eligen por un proceso aleatorio, como el uso de una tabla de números aleatorios.

Ensayo Clínico Controlado
Trabajo que consta de un ensayo clínico que implica uno o más tratamientos de prueba, al menos un tratamiento de control, medidas de control de resultados específicas para evaluar la intervención estudiada, y un método libre de sesgos para asignar los pacientes al grupo de prueba. El tratamiento puede ser medicamentos, aparatos o procedimientos analizados para eficacia diagnóstica, terapéutica, o profiláctica. Las medidas de control incluyen placebos, medicamentos activos, ningún tratamiento, formas y modalidades de dosis, comparaciones históricas, etc. Cuando la aleatorización utiliza técnicas matemáticas, como el uso de tabla de números aleatorios, para asignar pacientes al grupo de prueba o de control, se considera que es un 'Ensayo Controlado Aleatorizado'.

6.3.2.2 ¿Qué hay en el Registro Central Cochrane de Ensayos Controlados (CENTRAL) procedente de EMBASE?

En un estudio similar al descrito anteriormente para MEDLINE, se realizó una búsqueda en EMBASE por el Centro Cochrane del Reino Unido para informes de ensayos no indexados como ensayos en MEDLINE (Lefebvre 2008). (Los ensayos indexados como tales en MEDLINE ya están incluidos en CENTRAL, como se describe en la Sección 6.3.2.1, y por lo tanto no están duplicados en los registros EMBASE como parte del proceso de búsqueda). Los

siguientes términos son los usados corrientemente para el proyecto y han sido buscados desde 1980 hasta 2006: términos de texto libre:$aleator; $factorial; $cruzamient; $cruzar; $sobrepasar; $placebo; $dobl; $ciego adj; $ ciego adj $sencil; $asignac; $destin; $voluntar; y términos de índice, conocidos como términos EMTREE: procedimiento cruzado; procedimientos doble-ciego; ensayo controlado aleatorizado; procedimiento ciego-sencillo. Una búsqueda de los años 1974 a 1979 inclusive también se ha completado para los términos de texto libre: $aleator; $factori; $entrecruce y $placebo. El signo $ indica el uso de un símbolo truncado.
Estas búsquedas han producido un total de 80.000 informes de ensayos no indexados como informes de ensayos en MEDLINE, hasta el momento de la búsqueda. Todos estos registros están ahora publicados en CENTRAL, mediante contrato entre Elsevier, los editores de EMBASE, y la Colaboración Cochrane. De estos 80.000 registros, 50.000 son 'únicos' de CENTRAL, es decir, ya no están incluidos en CENTRAL con los registros obtenidos de MEDLINE. Esta búsqueda es actualizada anualmente. Las actualizaciones se describen en el archivo Detalles de Creación CENTRAL de *La Biblioteca Cochrane*:

- www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/CENTRALHelpFile.html

y la sección Qué es Nuevo de la página inicial de La Biblioteca Cochrane:

- www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME

6.3.2.3 ¿Qué hay en el Registro Central Cochrane de Ensayos Controlados (CENTRAL) procedente de otras bases de datos y de búsqueda manual?

Otras bases de datos de sanidad en general, como las publicadas en Australia y China, han realizado búsquedas sistemáticas similares para identificar informes de ensayos para CENTRAL. El Centro Cochrane Australiano-Asiático coordinó una búsqueda del Índice Médico Australiano-Asiático de la Biblioteca Nacional de Australia, desde 1966 (McDonald 2002). Esta búsqueda ha sido actualizada recientemente para incluir registros agregados hasta 2007. El Centro Cochrane de China, apoyado por el Centro Cochrane Australiano-Asiático, coordinó una búsqueda de la Base de Datos de la Literatura Biomédica China desde 1999 hasta 2001. En un proyecto en marcha, el Centro Cochrane de China, con el apoyo del Centro Cochrane del Reino Unido, está buscando varias fuentes chinas con miras a incluir estos registros en CENTRAL. Igualmente, el Centro Cochrane de Brasil, en colaboración con la Biblioteca Regional de Medicina en Brasil (BIREME), está planificando coordinar una búsqueda de la base de datos de la Organización Panamericana de la Salud (Literatura de Ciencias de la Salud del Caribe y Latinoamérica-LILACS).

Cada uno de los Centros Cochrane tiene la responsabilidad de buscar literatura general de sanidad en su país o región. Los CRGs y Áreas son los responsables de coordinar la búsqueda de la literatura sanitaria especializada en sus áreas de interés. Más de 3.000 revistas han sido, o están siendo, sometidas a búsqueda manual. Los informes de ensayos identificados que no estén relacionados con el área de un CGR y por lo tanto no son apropiados para su Registro especializado (ver abajo) son enviados a Wiley-Blackwell, a medida que haya resultados de esa búsqueda manual. Los registros de búsqueda manual pueden identificarse en CENTRAL porque están señalados con el rótulo HS-HANDSRCH o HS-PRECENTRL.

- www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/CENTRALHelpFile.html

6.3.2.4 ¿Qué hay en el Registro Central Cochrane de Ensayos Controlados (CENTRAL) procedente de los Registros Especializados de los Grupos y Áreas Cochrane de Revisión?

Es una 'función central esencial' de los CRGs que sus 'bases editoriales desarrollen y mantengan un Registro Especializado que contenga todos los estudios relevantes para su área de interés, y lo envíen a CENTRAL cada tres meses', como está definido en la Sección 3.2.1.5 'Funciones Centrales de los Grupos Cochrane de Revisión', en el Manual Cochrane. (www.cochrane.org/admin/manual.htm).

El Registro Especializado sirve para asegurar que los autores de revisión individual dentro del CRG tengan un acceso fácil y fiable a ensayos relevantes a su tema de revisión, normalmente a través de su Coordinador de Búsqueda de Ensayos. Los CRGs utilizan los métodos descritos en este Capítulo del *Manual* para identificar ensayos para sus registros especializados. La mayoría de los CRGs también tienen sistemas en el lugar para asegurar que cualquiera de los informes adicionales elegibles identificado por los autores para su(s) revisión(es) sea incluido en el Registro Especializado de su CRG. Los registros son, a su vez, enviados a CENTRAL para su inclusión, cada tres meses. De esta manera, los informes que se incluyen en el Registro Especializado de un CRG son accesibles a todos los otros CRGs a través de CENTRAL. Muchas Áreas también desarrollan Registros especializados de temas específicos y los envían para su inclusión en CENTRAL, como se describió antes. Para identificar informes en CENTRAL desde dentro de un Registro especializado específico es posible buscar en el rótulo Registro Especializado, tal como RE-Ictus. Una lista de todos los rótulos Registro Especializado puede encontrarse en el 'Apéndice: Grupo de Revisión/Códigos de la red de Registros Especializados' en la Carpeta de Ayuda 'Detalles de Creación Central' en *La Biblioteca Cochrane*: www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/CENTRALHelpFile.html

Informes en un Registro Especializado de un CRG a menudo contendrán códigos y otra información no incluida en CENTRAL, así el Coordinador de la Búsqueda de Ensayos a menudo podrá identificar informes adicionales en su Registro Especializado que podrían no ser identificados por búsqueda en CENTRAL, buscando estos códigos en el Registro Especializado. A la inversa, la funcionalidad de la búsqueda del software bibliográfico u otro utilizado para manejar los Registros especializados generalmente es menos sofisticado que la funcionalidad de la búsqueda disponible en La *Biblioteca Cochrane* así una búsqueda de CENTRAL recuperará informes del Registro Especializado que no son fácilmente recuperables desde dentro del Registro Especializado mismo. Se recomienda por lo tanto que CENTRAL y el Registro Especializado mismo sean consultados por separado para maximizar la recuperación.

6.3.3 Consultar CENTRAL, MEDLINE y EMBASE: asuntos específicos

Se recomienda que para todas las revisiones Cochrane se consulten CENTRAL y MEDLINE, como mínimo, junto con EMBASE si está disponible, bien en el CRG o para el autor de la revisión.

6.3.3.1 Consultar el Registro Central Cochrane de Ensayos Controlados (CENTRAL): asuntos específicos

CENTRAL se compone de registros de una amplia gama de fuentes (ver Sección 6.2.1.2 y 6.3.2 y subsecciones), de tal manera que no hay consistencia en el formato o en el contenido de los registros.

Los 310.000 registros procedentes de MEDLINE se recuperan mejor por una combinación de Títulos de Temas Médicos (MeSH, sigla en inglés) y términos de texto libre. Los otros registros, incluidos los 50.000 registros procedentes de EMBASE, se recuperan mejor utilizando búsquedas de texto libre en todas las áreas.

La mayoría de los registros que no proceden de MEDLINE o EMBASE (cerca de 170.000 en la *Biblioteca Cochrane* Volúmen 1, 2008) no tienen resúmenes o términos de indexación alguno. Para recuperar estos registros, que consisten predominantemente sólo en títulos, es necesario realizar una búsqueda muy amplia de una gama muy amplia de términos de texto libre, lo cual puede considerarse demasiado amplio para encontrarlos todos en CENTRAL.

Es posible identificar los registros que proceden de MEDLINE o EMBASE buscando en CENTRAL los registros que tengan número de acceso en PubMed o EMBASE. Es posible entonces excluir

estos registros de una búsqueda amplia en CENTRAL, como se muestra en el ejemplo del Recuadro 6.3.c.

Para información general acerca de la búsqueda, que tenga relación con la búsqueda en CENTRAL, ver Sección 6.4.

Recuadro 6.3.c: Ejemplo de exclusión de registros de MEDLINE y EMBASE cuando se consulta CENTRAL
Nota: el ejemplo tiene solo propósitos ilustrativos. Una búsqueda en CENTRAL para una revisión sistemática en este tema requeriría una escala amplia de términos alternativos para tamoxifeno y cáncer de mama.

#1	"número de acceso " near pubmed
#2	"número de acceso" near2 embase
#3	#1 o #2
#4	tamoxifeno
#5	(mama near cáncer)
#6	#4 y #5
#7	#6 no #3

6.3.3.2 Consultar MEDLINE y EMBASE: asuntos específicos

A pesar del hecho de que MEDLINE y EMBASE hayan sido consultados sistemáticamente para informes de ensayos, que estos informes de ensayos se han incluido en CENTRAL, como se describe en las Secciones 6.3.2.1 y 6.3.2.2, se recomiendan consultas adicionales en MEDLINE y EMBASE. Sin embargo, cualquiera de estas búsquedas debe emprenderse sabiendo lo que ya se ha hecho, para evitar duplicación de esfuerzos.

Consultar MEDLINE

Hay una demora de algunos meses entre que los registros sean indexados en MEDLINE y que aparezcan indexados como informes de ensayos en CENTRAL, ya que éste sólo se publica cada tres meses. Por ejemplo, para el volumen de *La Biblioteca Cochrane*, publicado en Enero de 2007, los registros MEDLINE fueron descargados por el personal de Wiley-Blackwell en Noviembre de 2006. La publicación de Enero de 2007 de *La Biblioteca Cochrane* fue el último volumen hasta Abril de 2007, de tal manera que los registros de MEDLINE están desactualizados entre dos y cinco meses. Lo meses más recientes de MEDLINE deberían por lo tanto consultarse al menos para registros indexados en los Tipos de Publicación bien cómo 'Ensayo Controlado Aleatorizado' o 'Ensayo Clínico Controlado', para identificar aquellos registros indexados como RCTs o CCTs en MEDLINE.
Adicionalmente, el año más reciente para ser consultado bajo el proyecto de identificar informes de ensayos en MEDLINE y enviarlos de regreso a la Biblioteca Nacional de Medicina de los Estados Unidos para ser rotulados de nuevo es el 2004, así que los registros agregados a MEDLINE durante y desde el 2005 deberían ser consultados utilizando una de las estrategias de búsqueda descritas en la Sección 6.4.11.1.

Finalmente, para sensibilidad extra, o donde el uso de un 'filtro' de ensayo aleatorizado no sea apropiado, los autores de la revisión deberían consultar MEDLINE para todos los años, utilizando sólo términos temáticos.

Debería recordarse que el proyecto de MEDLINE de rotular de nuevo descrito en la Sección 6.3.2.1 evaluó si los registros identificados fueron informes de ensayos en base sólo al título y el resumen, de tal manera que cualquier búsqueda adicional de MEDLINE que es continuada por el acceso al texto completo de los artículos identificará informes adicionales de ensayos, más probablemente a través de las secciones de métodos, que no fueron identificados solo por medio de los títulos o los resúmenes.

Cómo guía en realizar estrategias separadas de búsqueda en las versiones MEDLINE indexadas de MEDLINE y las versiones de MEDLINE que contienen registros 'en proceso' y otros registros no indexados, remitirse por favor a la Sección 6.4.11.1.
Cualquiera de los informes de ensayos identificados por el autor de la revisión puede ser remitido al Coordinador de Búsqueda de Ensayos, quien puede asegurar que serán incluidos en CENTRAL. Cualquiera de los errores respecto a registros indexados como ensayos en MEDLINE que con base al artículo completo definitivamente no son informes de ensayos de acuerdo a las definiciones utilizadas por la Biblioteca Nacional de Medicina (NLM, sigla en inglés) (ver Sección 6.3.2.1), debería también ser informado al Coordinador de Búsqueda de Ensayos, para que pueda remitirse a la NLM y ser corregido.

Para información general sobre consulta que esté relacionada con búsqueda en MEDLINE, ver la Sección 6.4.

Consultar EMBASE

El proyecto para identificar informes de ensayos en EMBASE para su inclusión en CENTRAL, descrito en la Sección 6.3.2.2, se realiza anualmente, así que hay un lapso de uno o dos años aproximadamente con respecto a los informes EMBASE que aparecen en CENTRAL. Los dos últimos años de EMBASE deberían por lo tanto ser buscados para cubrir trabajo aún en proceso. Algunos términos de búsqueda sugeridos se enumeran en la Sección 6.3.2.2. Un filtro de búsqueda diseñado por el equipo McMaster Hedges también está disponible (Wong 2006).

Finalmente, para sensibilidad extra, o donde no sea apropiado el uso de un 'filtro' de ensayo aleatorizado, los autores de la revisión deben consultar EMBASE para todos los años utilizando solo términos temáticos, como se describe arriba en circunstancias similares para MEDLINE. Debería recordarse que el proyecto EMBASE antes descrito evaluó si los registros identificados eran informes de ensayos con base solo en el título y el resumen, de igual manera que el proyecto MEDLINE descrito arriba. Por lo tanto, cualquier búsqueda adicional de EMBASE que es continuada por el acceso al texto completo de los artículos identificará informes adicionales de ensayos, lo más probable a través de las secciones de métodos, que no fueron identificados solo por medio de los títulos o los resúmenes.

Para infornación general sobre consulta, que se relacione con la consulta en EMBASE, ver la Sección 6.4.

6.3.4 Puntos de resumen

- Los autores de las revisiones Cochrane deberían pedir consejo a sus Coordinadores de Búsqueda de Ensayos a lo largo del proceso de búsqueda.
- Se recomienda que para todas las revisiones Cochrane deberían consultarse CENTRAL y MEDLINE, como mínimo, junto con EMBASE si está disponible bien para el CRG o para el autor de la revisión.
- Las estrategias completas para cada una de las bases de datos consultadas deberán incluirse en un Anexo de la revisión, de modo que todas las estrategias de búsqueda deberían guardarse, y anotarse el número de registros recuperados de cada una de las bases de datos consultadas.
- CENTRAL contiene cerca de 350.000 registros procedentes de MEDLINE y EMBASE, así que se debe tener cuidado al consultar MEDLINE y EMBASE para evitar duplicación innecesaria de trabajo.
- MEDLINE debería consultarse a partir de 2005 inclusive, utilizando una de las Estrategias Cochrane de Búsqueda Altamente Sensibles, revisada y actualizada, para identificar ensayos aleatorizados en MEDLINE como se estableció en la Sección 6.4.11.1.

- EMBASE debería consultarse para los dos últimos años como se estableció en la Sección 6.4.11.2.
- En MEDLINE y EMBASE se pueden identificar estudios adicionales buscando en los años ya consultados por CENTRAL, al obtener el artículo completo, y al leer, en particular, la sección de métodos.

6.4 Diseñar estrategias de búsqueda

6.4.1 Diseñar estrategias de búsqueda-una introducción

Esta sección resalta algunos de los asuntos a considerar cuando se diseñan estrategias de búsqueda, pero no muestra adecuadamente las múltiples complejidades de este área. Es precisamente en este aspecto de la búsqueda de estudios cuando más se recomiendan las habilidades de un Coordinador de Búsqueda de estudios o de un bibliotecario sanitario. Muchos de los asuntos destacados más adelante se relacionan tanto con el aspecto metodológico de la búsqueda (como identificar informes de ensayos aleatorizados) como con el tema de la búsqueda. Para que una búsqueda sea sólida los dos aspectos requieren igual atención, para estar seguros de no perder registros relevantes.

Los criterios de elegibilidad para incluir estudios en la revisión dirán cómo se realiza la búsqueda ver Capítulo 5). Los criterios de elegibilidad especificarán los tipos de diseños, los tipos de participantes, los tipos de intervención (experimental y comparativa) y, en algunos casos, los tipos de resultados que se deben buscar. Los aspectos a considerar al planificar una búsqueda incluyen los siguientes:

- si la revisión se limita a ensayos aleatorizados o si se incluirán otros diseños de estudio (también ver el Capítulo 13);
- el requisito para identificar datos de efectos adversos (ver también el Capítulo 14);
- la naturaleza de la(s) intervención(es) evaluada(s);
- algunas consideraciones geográficas como la necesidad de consultar la literatura China para los estudios de las plantas medicinales Chinas;
- el período de tiempo en que se realizan algunas evaluaciones de estas intervenciones; y
- si se van a incluir datos de estudios no publicados.

6.4.2 Estructura de una estrategia de búsqueda

La estructura de una estrategia de búsqueda debería basarse en los principales conceptos que se analizan en una revisión. Para una revisión Cochrane, el título de la revisión debería aportar estos conceptos, y los criterios de elegibilidad para incluir estudios ayudarán más en la selección de títulos temáticos apropiados y palabras de texto para la estrategia de búsqueda.

Generalmente es innecesario, y aún no deseado, consultar sobre cada aspecto de la pregunta clínica de la revisión (a menudo mencionada como PICO (sigla en inglés)-que se refiere a Paciente (o Participante o Población), Intervención, Comparación y Resultado). Aunque una pregunta de investigación puede referirse a poblaciones, escenarios o resultados particulares, estos conceptos pueden no estar bien descritos en el título o resumen de un artículo y a menudo no están bien indexados con términos de vocabulario controlado. Por lo tanto, generalmente no se prestan a una buena búsqueda. En bases de datos generales, como MEDLINE, una estrategia de búsqueda para identificar estudios para una revisión Cochrane tendrá clásicamente tres tipos de términos: 1) términos para buscar la condición de salud que interesa, es decir, la población; 2) términos para buscar la(s) intervención(es) evaluada(s); y 3) términos para buscar los tipos de diseño de estudios a incluir (típicamente un 'filtro' para ensayos aleatorizados). CENTRAL, sin embargo, busca contener sólo informes con diseños de

estudio posiblemente relevantes para ser incluidos en revisiones Cochrane, así que las búsquedas en CENTRAL no usarían un 'filtro' de ensayos. Los filtros para identificar ensayos aleatorizados y ensayos controlados se han desarrollado específicamente para MEDLINE y también se da una guía para buscar en EMBASE: ver Sección 6.4.11 y subsecciones. Para revisiones de intervenciones complejas puede ser necesario adoptar un enfoque diferente, por ejemplo, buscar sólo por la población o la intervención (Khan 2001).

6.4.3 Proveedores de servicio e interfases de búsqueda

Numerosos proveedores de servicio ofrecen tanto MEDLINE como EMBASE, mediante una escala de interfases de búsqueda; por ejemplo, Dialog ofrece Dialog y DataStar. Además la Biblioteca Nacional de Medicina de los Estados Unidos y Elsevier ofrecen acceso a sus propias versiones de MEDLINE y EMBASE respectivamente: MEDLINE a través de PubMed, que está disponible gratis en Internet, y EMBASE a través de EMBASE.com que está disponible sólo por suscripción. La sintaxis de búsqueda varía de una interfase a otra. Por ejemplo, para buscar el término Tipo de Publicación 'Ensayo Controlado Aleatorizado' en las diversas interfases de búsqueda es necesario introducir el término cómo:

Pt (tipo de publicación, iniciales en inglés). Ensayo controlado aleatorizado (en Ovid)
[pt] ensayo controlado aleatorizado (en PubMed)
Pt en ensayo controlado aleatorizado (en SilverPlatter)

Muchos proveedores de servicio ofrecen enlaces a versiones de texto completo de artículos en sitios web de otros editores, como la publicación 'Links/LinkOut' de PubMed.

6.4.4 Sensibilidad versus precisión

Las búsquedas para revisiones sistemáticas intentan ser tan extensas como puedan, con el fin de asegurar que se han incluido en la revisión tantos estudios necesarios y relevantes como sea posible. Sin embargo, es necesario hacer un balance entre esforzarse por la extensión y mantener la relevancia cuando se desarrolla una estrategia de búsqueda. Aumentando la extensión (o sensibilidad) de una búsqueda se reducirá su precisión y se recuperarán más artículos que no son relevantes.

Se define sensibilidad como el número de informes relevantes identificados dividido por el número total de informes relevantes existentes. La precisión es definida como el número de informes relevantes identificados dividido por el número total de informes identificados.

Desarrollar una estrategia de búsqueda es un proceso repetitivo en el que los términos que se usan son modificados, con base a lo que ya se ha recuperado. Las recompensas por los esfuerzos de búsqueda van disminuyendo; después de cierta etapa cada unidad adicional de tiempo invertido en la búsqueda aporta menos referencias que sean relevantes para la revisión. Como consecuencia se llega a un punto en el que las recompensas por una mayor búsqueda no compensan el esfuerzo requerido para identificar referencias adicionales. La decisión de cuánto invertir en el proceso de búsqueda depende de la pregunta que hace la revisión, la extensión en la que se desarrolla el Registro Especializado del CRG, y de los recursos disponibles. Debe anotarse, sin embargo, que los resúmenes de artículos identificados por medio de una búsqueda d e la literatura pueden tener una 'lectura escaneada' muy rápida para establecer su potencial relevancia. En un promedio de lectura, establecido de modo conservador, de dos resúmenes por minuto, los resultados de búsqueda en una base de datos pueden tener una 'lectura escaneada' de un promedio de 120 por hora (o aproximadamente 1.000 cada 8 horas), así que el alto rendimiento y la baja precisión asociados con la búsqueda de revisión sistemática o es tan desalentador como aparenta en comparación con el tiempo total invertido en la revisión.

6.4.5 Vocabulario controlado y palabras de texto

MEDLINE y EMBASE (y muchas otras bases de datos) pueden consultarse usando términos temáticos estandarizados asignados por indexadores. Los términos temáticos estandarizados (como parte de un vocabulario controlado o tesauro) son útiles porque pueden aportar una forma de recuperar artículos que pueden usar diferentes palabras para describir el mismo concepto y porque pueden aportar información más allá del simple contenido que tienen las palabras del título y del resumen. Sin embargo, cuando se buscan estudios para una revisión sistemática la extensión en la que se aplican los términos temáticos a las referencias debe verse con precaución. Los autores pueden no describir bien sus métodos u objetivos, y los indexadores no siempre son expertos en las áreas temáticas o en los aspectos metodológicos de los artículos que están indexando. Además, los términos indexados disponibles pueden no corresponder con los términos que el que consulta desea usar.

Los términos de búsqueda del vocabulario controlado para MEDLINE (MeSH) y EMBASE (EMTREE) no son idénticos, ni es el enfoque para indexar. Por ejemplo, los aspectos farmacéuticos o farmacológicos de un registro EMBASE son generalmente indexados en mayor profundidad que el registro equivalente en MEDLINE, y en los años recientes Elsevier ha aumentado el número de términos para indexar asignados a cada registro EMBASE. Las búsquedas de EMBASE pueden, por lo tanto, recuperar artículos adicionales que no fueron recuperados con una búsqueda MEDLINE, aún si los registros estaban en ambas bases de datos. Las estrategias de búsqueda necesitan ser adaptadas a cada base de datos.

Un modo de comenzar a identificar los términos del vocabulario controlado para una base datos determinada es recuperar artículos de esa base de datos que cumplan los criterios de inclusión para la revisión, y anotar palabras de texto comunes y los términos temáticos que los indexadores han aplicado a los artículos, que pueden entonces ser utilizados para una búsqueda completa. Habiendo identificado un artículo clave, se pueden localizar artículos relevantes adicionales, por ejemplo utilizando la opción 'Encontrar Similar' en Ovid o la opción 'Artículos Relacionados' en PubMed. Los términos adicionales de vocabulario controlado deberán identificarse utilizando las herramientas de búsqueda aportadas por la base de datos, tales como el Índice Permutado en Herramientas de Búsqueda, en Ovid, y la opción MeSH Database en PubMed.

Muchos tesauros de base de datos ofrecen la facilidad de 'explotar' términos temáticos para incluir más términos específicos automáticamente en la búsqueda. Por ejemplo, una búsqueda en MEDLINE que utiliza el término MeSH, LESIONES CEREBRALES, si se abre, automáticamente buscará no sólo el término LESIONES CEREBRALES sino también el término más específico SINDROME DEL NIÑO MALTRATADO. Como los artículos en MEDLINE en el tema de síndrome del niño maltratado deberían solamente ser indexados con el término más específico SÍNDROME DEL NIÑO MALTRATADO y no también con el término más general de LESIONES CEREBRALES es importante que los términos MeSH sean 'abiertos' donde sea más apropiado, con el fin de no perder artículos relevantes. El mismo principio se aplica a EMTREE cuando se consulta EMBASE y también a otra cantidad de bases de datos. Para más orientación en este tema, los autores de la revisión deben consultar a su Coordinador de Búsqueda de Ensayos o bibliotecario sanitario.

Es particularmente importante distinguir en MEDLINE entre los términos Tipo de Publicación y otros términos relacionados con MeSH. Por ejemplo, un informe de un ensayo aleatorizado deberá ser indexado en MEDLINE con el término Tipo de Publicación 'Ensayo Controlado Aleatorizado', mientras que un artículo sobre ensayos controlados aleatorizados deberá indexarse con el término MeSH ENSAYOS CONTROLADOS ALEATORIZADOS COMO TÓPICOS (notar que lo último es en plural). Lo mismo aplica para otros términos de indexación para ensayos, revisiones y metanálisis.

Los autores de la revisión deben asumir que los primeros artículos son más difíciles de identificar que los recientes. Por ejemplo, los resúmenes no están incluidos en MEDLINE para la mayoría de los artículos publicados antes de 1976 y, por tanto, las búsquedas por palabras de texto sólo se aplicarán a los títulos. Además, pocos términos de indexación MEDLINE

relacionados con el diseño del estudio estaban disponibles antes de 1990, así que las búsquedas de palabras de texto son necesarias para recuperar registros más antiguos.

Con el fin de identificar tantos registros relevantes como sea posible las búsquedas deberán comprender una combinación de términos temáticos seleccionados del vocabulario controlado o tesauros ('abierto' cuando sea apropiado) con un rango amplio de términos de texto libre.

6.4.6 Sinónimos, términos relacionados, variaciones ortográficas, abreviaturas y comodines

Cuando se diseña una estrategia de búsqueda, con el fin de que sea tan extensa como se pueda, es necesario incluir una gran cantidad de términos de texto libre para cada uno de los conceptos seleccionados. Por ejemplo:

- sinónimos: 'llaga de presión' OR 'úlcera de decúbito', etc;
- términos relacionados: 'cerebro' OR 'cabeza', etc; y
- variaciones ortográficas: 'tumour' OR 'tumor'.

Los proveedores de servicio ofrecen facilidades para capturar estas variaciones a través de abreviaturas y comodines:

- abreviatura: azar* (por azar o al azar, o aleatorio o aleatoriamente, etc); y
- comodín: mujer (es) (por mujer o mujeres).

Estos diseños pueden variar con los diferentes proveedores de servicios. Para mayores detalles consultar las carpetas de ayuda del proveedor del servicio de la base de datos correspondiente.

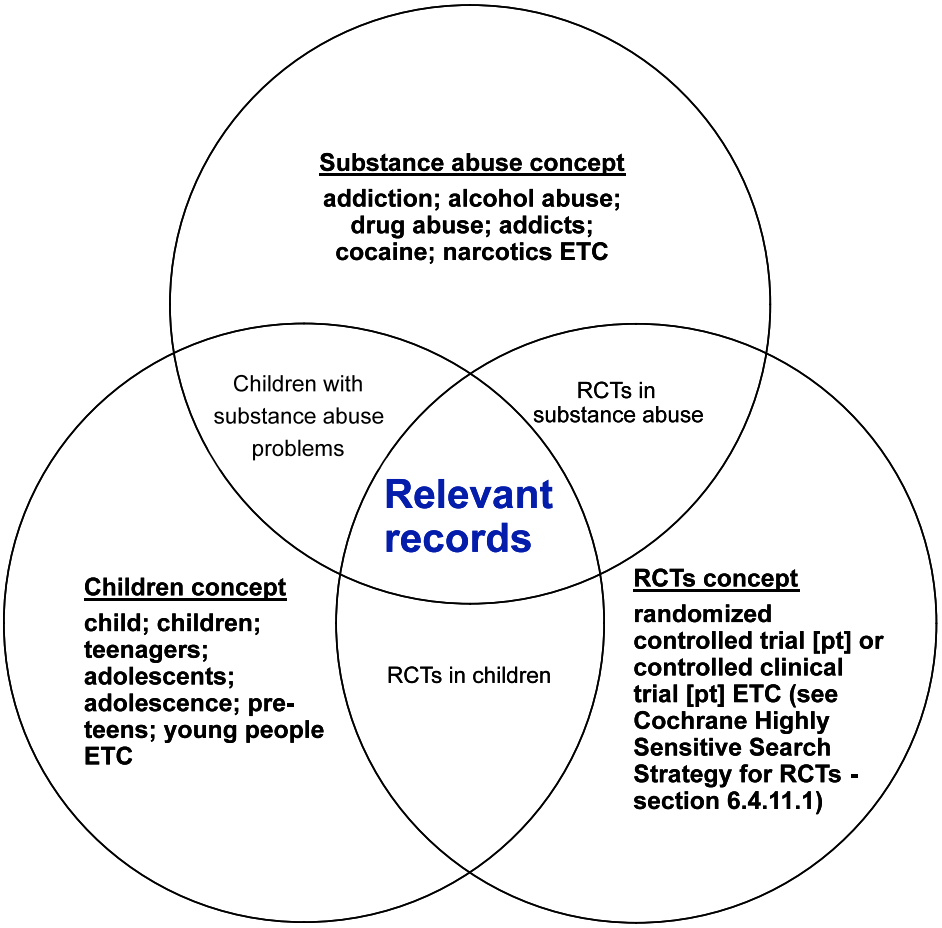
6.4.7 Los operadores Booleanos (AND, OR y NOT)

Una estrategia de búsqueda debería construir los términos de vocabulario controlado, las palabras de texto, los sinónimos y los términos relacionados para cada concepto de una vez, juntando cada uno de los términos dentro de cada concepto con el operador Booleano 'OR': ver estrategia de búsqueda (Figura 6.4.a). Esto significa que los artículos serán recuperados con al menos uno de estos términos de búsqueda. Las series de términos deberán por lo general ser elaboradas para la situación de salud, la(s) intervención(es) y el diseño del estudio. Estas tres series de términos pueden ser juntadas con el operador 'AND'. Este paso final de juntar las tres series con el operador 'AND' limita el conjunto recuperado a artículos del diseño de estudio apropiado que se refiere tanto a la condición de salud que interesa como a la(s) intervención(es) que se va a evaluar. Sin embargo, se justifica una nota de precaución acerca de este enfoque: si un artículo no contiene al menos un término de cada una de las tres series, no será identificado. Por ejemplo, si un término del índice no se ha agregado al registro de la intervención y la intervención no es mencionada en el título y en el resumen, el artículo se perdería. Una posible solución es omitir una de las tres series de términos y decidir qué registros revisar en base al número recuperado y el tiempo disponible para revisarlos. El operador 'NOT' debería eludirse cuando sea posible para evitar el peligro de eliminar inadvertidamente de la serie de búsqueda registros que son relevantes. Por ejemplo, cuando se buscan registros indexados como femenino, el 'masculino NOT' eliminaría cualquier registro que fuera sobre masculinos y femeninos.

Las búsquedas para las revisiones Cochrane pueden ser extremadamente largas y con frecuencia incluyen cerca de 100 declaraciones de búsqueda. Puede ser tedioso escribir las combinaciones de estas series de búsqueda, por ejemplo como '#1 OR #2 OR #3 OR #4….OR #100'. Algunos proveedores de servicio ofrecen alternativas a esto. Por ejemplo, en Ovid es posible combinar series utilizando la sintaxis 'or/1-100'. Para aquellos proveedores de servicio donde esto no es posible, incluida *La Biblioteca Cochrane* para búsquedas de CENTRAL, se ha recomendado que la búsqueda mencionada arriba podría ser escrita completa y guardada, por ejemplo, como un documento Word y el número requerido de combinaciones copiadas y

pegadas dentro de la búsqueda de acuerdo a la necesidad. Habiendo escrito la serie con los símbolos # como arriba, se puede generar una segunda serie al reemplazar totalmente el símbolo # por nada para crear la serie '1 OR 2 OR 3 OR 4 .... OR100, para usarse en estos proveedores de servicio donde la interfase de búsqueda no use el símbolo#.

Figura 6.4.a: Combinar conceptos como grupos de búsqueda



6.4.8 Operadores cercanos (NEAR, NEXT y ADJ)

En algunas interfases de búsqueda es necesario especificar, por ejemplo, utilizando el operador 'NEXT' o 'ADJ' que dos términos de búsqueda deberían ser adyacentes el uno del otro, porque la búsqueda podría fallar simplemente al no encontrar ambas palabras en el documento como si el operador 'AND' hubiera sido utilizado. Deberá notarse que el operador 'NEXT' en *La Biblioteca Cochrane* es más sensible (es decir, recupera más) que el método alternativo de la frase de búsqueda que usa signos de puntuación, ya que los signos de puntuación especifican esa frase mientras el operador 'NEXT' incorpora auto-pluralización y auto-singularización tanto como otras terminaciones variables de palabras.

Además, es posible en muchas interfases de búsqueda especificar que las palabras deberían estar dentro de un número específico de palabras una de la otra. Por ejemplo, el operador 'NEAR' en *La Biblioteca Cochrane* encontrará los términos de búsqueda dentro de seis palabras la una de la otra. Esto resulta en una sensibilidad más alta que buscar por una frase sencilla o el uso del operador 'NEXT', pero con mayor precisión que el uso del operador 'AND'. Por lo tanto, es deseable usar este operador cuando esté disponible y sea relevante.

6.4.9 Lenguaje, fecha y restricciones del formato del documento

La investigación relacionada con la identificación de ensayos se ha centrado recientemente en el efecto de excluir frente a incluir ensayos de metanálisis informados en idiomas diferentes al inglés. Este asunto es particularmente importante porque la identificación y traducción, o al menos la extracción de datos de los ensayos informados en idiomas distintos del inglés pueden recargar de manera importante los costes de una revisión, y el tiempo que toma el completarla. Para mayor información sobre estos asuntos, ver el Capítulo 10 (Sección 10.2.2.4). En la medida de lo posible, los autores de la revisión deberían intentar identificar y evaluar para elegibilidad todos los informes de ensayos posiblemente relevantes, independientemente del idioma de su publicación. No deberían incluirse restricciones de idioma en la estrategia de búsqueda. Las restricciones de fecha deberían aplicarse si se sabe que los estudios relevantes sólo podrían haber sido informados durante un período de tiempo específico, por ejemplo si la intervención sólo estuvo disponible después de cierto tiempo. Restricciones de formato como excluir cartas no se recomiendan porque las cartas pueden contener información adicional importante relacionada con un informe de ensayo más anterior, o información nueva de un ensayo no informado en otra parte.

6.4.10 Identificar estudios falsos, otras publicaciones retiradas, errores de imprenta y comentarios

Al considerar la elegibilidad de estudios para su inclusión en una revisión Cochrane, es importante vigilar porque se ha encontrado que algunos estudios son falsos, o por varias otras razones han sido retirados después de su publicación. Los informes de estudios indexados en MEDLINE que han sido retirados (por ser falsos o por otras razones) tendrán el término Tipo de Publicación 'Publicación Retirada' agregado al registro. El artículo que da la noticia del retiro tendrá asignado el término Tipo de Publicación 'Retiro de Publicación'. Antes de tomar la decisión de retirar un artículo, pueden publicarse artículos que se refieren a un artículo original y aumentar las inquietudes correspondientes. Tales artículos deberían clasificarse como Comentario. La política de la Biblioteca Nacional de Medicina de los Estados Unidos (NLM) sobre esto es que "Entre los tipos de artículos que serán considerados comentarios están: ..... anuncios o noticias que informan de ciencia cuestionable o investigaciones por mala conducta científica (algunas veces publicadas como 'Expresión de inquietud')".

- www.nlm.nih.gov/pubs/factsheets/errata.html

Además, los artículos pueden haber sido parcialmente retirados, corregidos por medio de una fe de errata o pueden haber sido corregidos y vueltos a publicar por completo. Al actualizar una revisión, es importante buscar en MEDLINE las últimas versiones de las citas a los registros de los estudios incluidos. En ciertos formatos de algunas versiones de MEDLINE las declaraciones de la publicación retirada, la fe de erratas y los comentarios se incluyen en la cita inmediatamente después del título y son, por lo tanto, bien visibles. Sin embargo, este no es siempre el caso, por lo cual se debe tener cuidado de asegurarse de que esta información siempre se recupera en todas las búsquedas al descargar los campos apropiados junto con los datos de las citas (ver Sección 6.5.2). Para mayores detalles de la política y la práctica de la NLM en este tema ver:

- www.nlm.nih.gov/pubs/factsheets/errata.html

6.4.11 Filtros de búsqueda

Los filtros de búsqueda son estrategias de búsqueda diseñadas para recuperar tipos específicos de registros, como los de un diseño metodológico particular. Pueden ser estrategias derivadas subjetivamente como la original Estrategia de Búsqueda Cochrane de Alta Sensibilidad, para identificar informes de ensayos aleatorizados en MEDLINE) (Dickersin

1994) o pueden ser objetivamente derivadas por el análisis de frecuencia de palabra y probadas en series de datos de registros relevantes para evaluar su sensibilidad y precisión, como las estrategias de búsqueda que se detallan abajo para identificar ensayos aleatorizados en MEDLINE (Glanville 2006). Recientemente un sitio web de filtros de búsqueda ha sido desarrollado por el Subgrupo de Especialistas en Información InterTASC del Reino Unido (ISSG, sigla en inglés), que es el grupo de profesionales de la información que apoyan a grupos de investigación dentro de Inglaterra y Escocia, aportando evaluaciones tecnológicas al Instituto Nacional para la Salud y la Excelencia Clínica (NICE) (Glanville 2008). El propósito del sitio web es enumerar los filtros de búsqueda metodológica y hacer valoraciones críticas de los diversos filtros. El sitio incluye, entre otros, filtros para identificar revisiones sistemáticas, estudios aleatorizados y no aleatorizados e investigación cualitativa en un conjunto de bases de datos y por medio de un conjunto de proveedores de servicio.

- www.york.ac.uk/inst/crd/intertasc/

Los filtros de búsqueda deben usarse con precaución. Deberían evaluarse no sólo por la fiabilidad de su desarrollo y su actuación informada sino también por su precisión corriente, su relevancia y efectividad, dada la frecuente interfase y los cambios de indexación que afectan a las bases de datos.
El ISSG ofrece una herramienta de valoración del filtro de búsqueda para ayudar a evaluar los filtros de búsqueda y los ejemplos pueden verse en el sitio web.

- www.york.ac.uk/inst/crd/intertasc/qualitat.htm

6.4.11.1 Las estrategias Cochrane de Búsqueda Altamente Sensible para identificar ensayos aleatorizados en MEDLINE

La primera estrategia Cochrane de Búsqueda Altamente Sensible para identificar ensayos aleatorizados en MEDLINE fue diseñada por Carol Lefebvre y publicada en 1994 (Dickersin 1994). Esta estrategia fue publicada posteriormente en el *Manual* y ha sido adaptada y actualizada por necesidad a lo largo del tiempo. Las Estrategias Cochrane de Búsqueda Altamente Sensible para MEDLINE en las secciones siguientes son adaptadas de las estrategias publicadas primero en 2006, como resultado de un análisis de frecuencia de términos MeSH y términos de texto libre realizado en los títulos y resúmenes de los registros de informes de ensayos aleatorizados indexados en MEDLINE (Glanville 2006), utilizando métodos de diseño de estrategia de búsqueda desarrollados primero por los autores para identificar revisiones sistemáticas en MEDLINE (White 2001).

Se ofrecen dos versiones: una versión que maximiza la sensibilidad y una versión que maximiza la sensibilidad y la precisión. Se recomienda que las búsquedas de ensayos para su inclusión en las revisiones Cochrane comiencen con la versión que maximiza la sensibilidad, en combinación con una búsqueda temática altamente sensible. Si esta versión recupera un número inmanejable de referencias, en ese caso debería utilizarse la versión que maximiza la sensibilidad y la precisión. Debería tenerse en mente que los resúmenes de MEDLINE pueden leerse muy rápidamente porque son relativamente cortos, y en un cálculo cauteloso de 30 segundos por resumen, en 8 horas se pueden leer aproximadamente 1.000 resúmenes.

Las estrategias han sido actualizadas después de re-analizar los datos utilizados para obtener esas estrategias, para reflejar los cambios en la política de indexación introducida por la Biblioteca Nacional de Medicina de los Estados Unidos, diferentes de los análisis y cambios originales en la sintaxis de búsqueda. Estos cambios incluyen:

- no seguir asignando ‘Ensayo Clínico’ como un Tipo de Publicación a todos los registros indexados como ‘Ensayo Controlado Aleatorizado’ o ‘Ensayo Clínico Controlado’ bajo Tipo de Publicación; y
- el cambio del término MeSH ENSAYOS CLÍNICOS a ENSAYOS CLÍNICOS COMO TEMA.

Las estrategias se encuentran en el Recuadro 6.4.a y en el Recuadro 6.4.b para PubMed, y en el Recuadro 6.4.c y en el Recuadro 6.4.d para Ovid.

Debe tenerse en mente que las estrategias que siguen están basadas en datos derivados de los registros indexados en MEDLINE y están diseñadas para ser consultadas en MEDLINE. Estas estrategias no están diseñadas para recuperar registros 'en proceso' y otros no indexados en MeSH. Por lo tanto se recomienda que estas estrategias sean consultadas en MEDLINE – las versiones indexadas de MEDLINE y las búsquedas separadas para registros no indexados se consultan en la base de datos que contiene el 'en proceso' y registros no indexados. Por ejemplo, en Ovid las estrategias de abajo deberían consultarse y actualizarse en bases de datos como 'Ovid MEDLINE (R)1950 a Mes SemanaX 200X' y los registros no indexados deberían buscarse en 'Ovid MEDLINE (R) En Proceso & Otras Citas No Indexadas MesX, 200X'. Para identificar registros no indexados se requeriría un rango de términos de texto libre abreviados, tales como aleatorio, placebo, ensayo, etc, y la búsqueda no se debe limitar a humanos (ya que los registros no están todavía indexados como humanos).

Como se mencionó en la Sección 6.3.2.1, MEDLINE ha sido registrado desde 1966 hasta 2004 inclusive, utilizando versiones previas de la Estrategia Cochrane de Búsqueda Altamente Sensible, paras identificar ensayos aleatorizados, y los registros de informes de ensayos se han re-indexado en MEDLINE e incluidos en CENTRAL (solamente en base a los títulos y los resúmenes). Ir a la Sección 6.3.2.1 y 6.3.3.2 para una mayor información en el uso apropiado de estas Estrategias de Búsqueda Altamente Sensible.

Recuadro 6.4.a: Estrategia Cochrane de Búsqueda Altamente Sensible para identificar ensayos aleatorizados en MEDLINE: versión de máxima sensibilidad (revisada en 2008); formato PubMed

#1	ensayo controlado aleatorizado [pt]
#2	ensayo clínico controlado [pt]
#3	aleatorizado [tiab]
#4	placebo [tiab]
#5	terapia con fármaco [sh]
#6	aleatoriamente [tiab]
#7	ensayo [tiab]
#8	grupos [tiab]
#9	#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8
#10	animales [mh] not (humanos [mh] and animales [mh])
#11	#9 not #10

Sintaxis de búsqueda de PubMed
[pt] denota un término Tipo de Publicación;
[tiab] denota una palabra en el título o en el resumen;
[sh] denota un subtítulo;
[mh] denota un término de Título de Tema Médico (MeSH) ('explotado');
[mesh: noexp] denota un término de Título de Tema Médico (MeSH) (no 'explotado');
[ti] denota una palabra en el título.

Recuadro 6.4.b: Estrategia Cochrane de Búsqueda Altamente Sensible para identificar ensayos aleatorizados en MEDLINE: versión de máxima sensibilidad y precisión (revisada en 2008); formato PubMed

#1	ensayo controlado aleatorizado [pt]
#2	ensayo clínico controlado [pt]
#3	aleatorizado [tiab]
#4	placebo [tiab]
#5	ensayos clínicos como tema [mesh: noexp]
#6	aleatoriamente[tiab]
#7	ensayo [ti]
#8	#1 or #2 or #3 or #4 or #5 or #6 or #7
#9	animales [mh] not (humanos [mh] and animales [mh])
#10	#8 not #9

La sintaxis de búsqueda se explica en el Recuadro 6.4.a.

Recuadro 6.4.c: Estrategia Cochrane de Búsqueda Altamente Sensible para identificar ensayos aleatorizados en MEDLINE: versión de máxima sensibilidad (revisada en 2.008); formato Ovid

1	ensayo controlado aleatorizado.pt.
2	ensayo clínico controlado.pt.
3	aleatorizado.ab.
4	placebo.ab.
5	farmacoterapia.fs.
6	aleatoriamente.ab.
7	ensayo.ab.
8	grupos.ab.
9	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10	animales.sh. not (humanos.sh. and animales.sh.)
11	9 not 10

Sintaxis de búsqueda de Ovid
.pt. denota un término de Tipo de Publicación;
.ab. denota una palabra en el resumen;
.fs. denota un subtítulo 'flotante';
.sh. denota un término de Título de Tema Médico (MeSH);
.ti. denota una palabra en el título.

Recuadro 6.4.d: Estrategia Cochrane de Búsqueda Altamente Sensible para identificar ensayos aleatorizados en MEDLINE: versión de máxima sensibilidad y precisión (revisada en 2.008); Formato de Ovid

1	ensayo controlado aleatorizado.pt.
2	ensayo clínico controlado.pt.
3	aleatorizado.ab.
4	placebo.ab.
5	ensayos clínicos como tema.sh.
6	aleatoriamente.ab.
7	ensayo.ti.
8	1 or 2 or 3 or 4 or 5 or 6 or 7
9	animales.sh. not (humanos.sh. and animales.sh.)
10	9 not 10

La sintaxis de búsqueda se explica en el Recuadro 6.4.c.

6.4.11.2 Filtros de búsqueda para identificar ensayos aleatorizados en EMBASE

El Centro Cochrane del Reino Unido está trabajando en el diseño de una estrategia de búsqueda altamente sensible derivada objetivamente para identificar informes de ensayos aleatorizados en EMBASE, utilizando métodos de análisis de frecuencia de palabra similares a los utilizados para diseñar las estrategias de búsqueda altamente sensibles para identificar ensayos aleatorizados en MEDLINE descritos en la Sección 6.4.11.1 (Glanville 2006). Los autores de revisión que deseen dirigir sus propias búsquedas de EMBASE mientras tanto podrían considerar el uso de los términos de búsqueda enumerados en la Sección 6.3.2.2 que son los usados normalmente por el Centro Cochrane del Reino Unido para identificar informes de EMBASE de ensayos aleatorizados para su inclusión en CENTRAL (Lefebvre 2008). Alternativamente, el filtro de búsqueda diseñado por Wong y cols. para identificar lo que ellos denominan ”estudios que parecen clínicamente sólidos” en EMBASE puede utilizarse (Wong 2006).

Como se mencionó en la Sección 6.3.2.2, EMBASE ha sido registrado desde 1.980 hasta 2.006 inclusive, utilizando los términos enumerados en esa sección, y los registros de los informes de ensayos (basados en los títulos y los resúmenes solamente) se han incluido en CENTRAL.

6.4.12 Actualizar las búsquedas

Cuando se actualiza una revisión Cochrane el proceso de búsqueda tendrá que ser revisado (es decir, decidir qué bases de datos y otras fuentes de búsqueda para cuales años). Esas bases de datos que previamente fueron registradas y se consideran relevantes para la actualización necesitan ser registradas de nuevo. Las estrategias de búsqueda previas necesitarán ser actualizadas para reflejar aspectos como: cambios en la indexación tales como la adición o retiro de términos de vocabulario controlado (MeSH, EMTREE, etc); cambios en la sintaxis de búsqueda; comentarios o críticas sobre las estrategias de búsqueda previas. Si alguna de las bases de datos inicialmente registrada no lo será para la actualización, esto debe ser explicado y justificado. Nuevas bases de datos u otras fuentes pueden haberse producido o hacerse disponibles para el autor de la revisión o para el Coordinador de la Búsqueda de Ensayos y estas también deberían ser consideradas.

Se debería tener cuidado con el uso de los límites de actualización cuando se busca por medio de registros indexados y no indexados en MEDLINE simultáneamente, tales como en PubMed o en el archivo de Ovid MEDLINE ‘En Proceso & Otras Citas no Indexadas y Ovid MEDLINE de 1950 hasta el Presente’. Cuando sea posible, los archivos separados deberían ser seleccionados y consultados de forma separada, como el archivo ‘1950 a Semana Mes X 200X’ de Ovid MEDLINE, y los registros no indexados deberían buscarse en el archivo ‘En Proceso & Otras Citas No Indexadas MesX 200X’ de Ovid MEDLINE. Para más guía en este tema contactar con los Coordinadores de Búsqueda de Ensayos.

6.4.13 Demostración de las estrategias de búsqueda

El Recuadro 6.4.e hace una demostración de estrategia de búsqueda para CENTRAL del tema ‘Tamoxifeno para el cáncer de mama’. Notar que solo incluye términos temáticos (un filtro para un ensayo aleatorizado no es apropiado para CENTRAL). No está limitado solo a humanos. La estrategia se presenta sólo con propósitos ilustrativos: las búsquedas de CENTRAL para estudios a incluir en una revisión sistemática debería tener muchos más términos de búsqueda para cada uno de los conceptos
El Recuadro 6.4.f hace una demostración de estrategia de búsqueda para MEDLINE (formato Ovid) para el tema ‘Tamoxifeno para cáncer de mama’. Notar que se utilizan tanto términos temáticos como filtro para ensayo aleatorizado en MEDLINE. La búsqueda está limitada a humanos. La estrategia se presenta sólo con propósitos ilustrativos: las búsquedas de

MEDLINE para revisiones sistemáticas debería tener muchos más términos de búsqueda para cada uno de los conceptos.

Recuadro 6.4.e: Demostración de la estrategia de búsqueda para CENTRAL, para el tema 'Tamoxifeno para cáncer de mama'

#1	El descriptor MeSH Neoplasmas de Mama explora todos los árboles
#2	mama near cáncer *
#3	mama near neoplasma*
#4	mama near carcinoma*
#5	mama near tumour*
#6	mama near tumor*
#7	#1 OR #2 OR #3 OR #4 OR #5 OR #6
#8	El descriptor MeSH Tamoxifeno explora todos los árboles
#9	tamoxifeno
#10	#8 OR #9
#11	#7 AND #10

El operador 'near' toma por defecto hasta seis palabras;
'*' indica abreviatura.

Recuadro 6.4.f: Demostración de estrategia de búsqueda para MEDLINE (formato Ovid), en el tema 'Tamoxifeno para cáncer de mama'

1	Ensayo controlado aleatorizado.pt.
2	Ensayo clínico controlado.pt.
3	aleatorizado.ab.
4	placebo.ab.
5	farmacoterapia.fs.
6	aleatoriamente.ab.
7	ensayo.ab.
8	groupos.ab.
9	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10	animales.sh. not (humanos.sh. and animales.sh.)
11.	9 not 10
12.	exp Neoplasmas de Mama/
13.	(mama adj6 cáncer$).mp.
14.	(cáncer adj6 neoplasm$).mp.
15.	(mama adj6 carcinoma$).mp.
16.	(mama adj6 tumour$).mp.
17.	(mama adj6 tumor$).mp.
18.	12 or 13 or 14 or 15 or 16 or 17
19.	exp Tamoxifeno/
20.	tamoxifeno.mp.
21.	19 or 20
22.	11 and 18 and 21

El operador 'adj6' indica hasta seis palabras;
'$' indica abreviatura;
.mp. indica una búsqueda de título, título original, resumen, nombre de la palabra esencial y la palabra del título del tema.

6.4.14 Puntos de resumen

- Los autores de revisiones Cochrane deberían contactar con sus Coordinadores de Búsqueda de Ensayos *antes* de comenzar la búsqueda.

- Para la mayoría de las revisiones Cochrane, la estructura de búsqueda en la mayoría de las bases de datos estará constituida por una búsqueda temática para la población o la condición y la intervención junto con un filtro metodológico para el diseño del estudio, como los ensayos aleatorizados.
- Para las búsquedas en CENTRAL, no aplicar un filtro de ensayo aleatorizado y no limitarla a lo humano.
- Evitar demasiados conceptos de búsqueda diferentes pero usar una variedad amplia de términos sinónimos y relacionados (tanto términos de texto libre como de vocabulario controlado) combinados con 'OR' dentro de *cada* concepto.
- Combinar conceptos diferentes con 'AND'.
- Evitar el uso del operador 'NOT' en combinación con series de búsqueda.
- Pretender alta sensibilidad y estar preparado para aceptar baja precisión.
- No aplicar restricciones lingüísticas a la estrategia de búsqueda.
- Las búsquedas diseñadas para una base de datos y un proveedor de servicio específicos necesitarán 'traducirse' para usarse en otra base de datos o en otro proveedor.
- Estar alerta a cualquiera de las publicaciones retiradas (por ejemplo publicaciones falsas), fe de erratas y comentarios.
- Para identificar ensayos alatorizados en MEDLINE comience con la versión que maximiza la sensibilidad, la Estrategia Cochrane de Búsqueda Altamente Sensible. Si esta recupera un número inmanejable de referencias, utilizar en su lugar la versión que maximiza la sensibilidad y la precisión.
- Para actualizar las búsquedas, en lo posible, los archivos de bases de datos separadas deberían ser seleccionados y consultados por aparte de los registros indexados y no indexados en proceso de MEDLINE.

6.5 El manejo de las referencias

6.5.1 El software bibliográfico

El software especialmente diseñado para el manejo bibliográfico o de referencias como EndNote, ProCite, Reference Manager y RefWorks es útil y relativamente fácil de usar para llevar la cuenta de las referencias y de los informes de los estudios. La elección de cuál software utilizar probablemente está influenciada por lo que esté disponible y por tanto tenga el apoyo de la institución del autor de la revisión. Para una comparación de los productos anteriores y enlaces a revisiones de otros paquetes de software bibliográfico ver:

- www.burioni.it/forum/dellorso/bms-dasp/text/

De los paquetes enumerados arriba, ProCite generalmente está considerado como muy eficiente para identificar referencias duplicadas, pero sus proveedores no lo han actualizado. No apoya el rango más amplio de conjuntos de caracteres que permiten introducir correctamente idiomas diferentes del inglés, mientras que Endnote sí lo hace. El software bibliográfico también facilita el almacenamiento de datos acerca de los métodos y de los procesos de una búsqueda. Por ejemplo, los campos separados no usados pueden utilizarse para almacenar información como: 1) el nombre de la base de datos o detalles de otra fuente de la cual se identificó el informe de un ensayo, 2) cuándo y desde dónde se ordenó un artículo y la fecha de recibo del artículo y 3) si el estudio asociado con un artículo fue incluido o excluido de una revisión, y si fue excluido, las razones de su exclusión.

Los archivos para importar referencias desde CENTRAL en el software bibliográfico están disponibles en el sitio web del Grupo Cochrane de Métodos de Recuperación de Información en:

- www.cochrane.org/docs/import.htm

6.5.2 Qué campos descargar

Además de la cita completa del registro, una cantidad de campos clave podrían considerarse para descargar de las bases de datos cuando estén disponibles. Una guía más detallada de qué campos descargar ha sido recogida por el Grupo de Trabajo de los Coordinadores de la Búsqueda de Ensayos y está disponible en un documento titulado 'Guía de Usuarios para el Manejo Especializado de Registros y de Registros de Búsqueda Manual, de los Coordinadores de Búsqueda de Ensayos', en:

- www.cochrane.org/resources/hsearch.htm

Resumen: los resúmenes pueden usarse para eliminar de manera evidente informes irrelevantes, para obviar la necesidad de obtener el texto completo de esos artículos, o volver más tarde a la base de datos bibliográfica.

Número de acceso/identificador único: se aconseja apartar un campo no utilizado para almacenar el identificador único/el número de acceso de los registros descargados, como el número ID de PubMed (PMID). Esto permite un vínculo posterior al registro completo de la base de datos y también facilita el manejo de información tal como la detección y cancelación de duplicados.

Afiliación/dirección: puede incluir la afiliación institucional y/o la dirección electrónica del (los) autor(es).

Identificador de artículos/identificador de objeto digital (DOI, sigla en inglés): puede utilizarse para citar y vincular el registro completo.

Número de ensayo clínico: si el registro contiene un número de ensayo clínico como los asignados por el Ensayos Clínicos.gov o esquemas ISRCTN o un número asignado por el patrocinador del ensayo, estos deberán ser descargados para ayudar a la vinculación de los informes del ensayo con los estudios originales. Un ejemplo de esto es el Número de Ensayo Clínico (CN), campo introducido recientemente en EMBASE.

Términos del índice/términos tesauros/palabras clave: ver la Sección 6.4.5. Estos ayudan a indicar porqué los registros fueron recuperados si el título y el resumen carecen de detalle.

Lenguaje: lenguaje de publicación del artículo original.

Comentarios, correcciones, errores de impresión, retractaciones y actualizaciones: es importante asegurar que cualquiera de los campos que se relacionen con comentarios, correcciones, errores de impresión, retractaciones y actualizaciones publicados posteriormente se seleccionen para su inclusión en la descarga, de manera que cualquier impacto de estas publicaciones posteriores puedan tomarse en cuenta. Los campos más importantes a considerar, junto con sus etiquetas de campo en PubMed, se presentan en el Recuadro 6.5.a.

- www.nlm.nih.gov/bsd/mms/medlineelements.html#cc

Recuadro 6.5.a: Etiquetas de campo importantes en PubMed

CIN: 'Comentar en'
CON: 'Comentario sobre'
CRI: 'Corregido y publicado de nuevo en'
CRF: ''Corregido y publicado de nuevo desde'
EIN: 'Error en'
EFR: 'Error por'
PRIN: 'Retractación parcial en'
PROF: 'Retractación parcial de'
RIN: 'Retractación en'
ROF: 'Retractación de'
RPI: 'Publicado de nuevo en'
RPF: ''Publicado de Nuevo desde'
UIN: 'Actualizado en'
UOF: 'Actualizado de'

6.5.3 Puntos de resumen

- Usar software bibliográfico para manejar las referencias.
- Asegurarse de que todos los campos necesarios sean descargados.

6.6 Documentar e informar el proceso de búsqueda

6.6.1 Documentar el proceso de búsqueda

El proceso de búsqueda debe documentarse con suficiente detalle a lo largo del proceso para asegurar que puede ser informado correctamente en la revisión, hasta el punto de que todas las búsquedas en todas las bases de datos puedan ser reproducidas. Debería tenerse en cuenta desde el principio que las estrategias de búsqueda completas para cada base de datos necesitarán incluirse en un Anexo de la revisión. Las estrategias de búsqueda deberán copiarse exactamente como se llevaron a cabo e incluidas en su totalidad, junto con los números de serie de la búsqueda y el número de registros recuperados. El número de registros recuperados deberá ser registrado en la sección de Resultados de la revisión, bajo el título 'Resultados de la búsqueda' (ver Capítulo 4, Sección 4.5). Las estrategias de búsqueda no deben ser escritas de nuevo porque esto puede introducir errores. Un estudio reciente ha mostrado falta de cumplimiento con lo que recomienda el *Manual* con respecto a la descripción de la estrategia de búsqueda en las revisiones Cochrane (Sampson 2006). En la mayoría de los CRGs, a los Coordinadores de Búsqueda de Ensayos ahora se les ha pedido que comenten sobre las secciones de estrategia de búsqueda de una revisión como parte del proceso de finalización, antes de que una revisión se considere lista para su publicación en el *CDSR*. Por lo tanto se recomienda que los autores de la revisión busquen la orientación de sus Coordinadores de Búsqueda de Ensayos a la primera oportunidad, con respecto a documentar el proceso para facilitar la redacción de esta sección de la revisión. Como se mencionó en alguna parte de este Capítulo, es particularmente importante guardar enseguida o imprimir copias del archivo de cualquier información que se encuentre en Internet, como la información de los ensayos en marcha, porque esta información puede que no esté disponible cuando se finalice la revisión.

6.6.2 Informar el proceso de búsqueda

6.6.2.1 Informar el proceso de búsqueda en el protocolo

La inclusión de cualquiera de las estrategias de búsqueda en el protocolo de una revisión Cochrane es opcional. Donde las búsquedas ya han sido realizadas en la fase del protocolo se considera útil incluirlas en el protocolo. Así pueden ser comentadas de la misma manera que otros aspectos del protocolo. Algunos CRGs piensan que las búsquedas no deberían comenzar hasta que el protocolo esté terminado para su publicación porque el conocimiento de los estudios disponibles podrían influenciar aspectos del mismo, como los criterios de inclusión.

6.6.2.2 Informando el proceso de búsqueda en la revisión

Informando el proceso de búsqueda en el resumen de la revisión

- Enumere todas las bases de datos consultadas.
- Anote las fechas de la última búsqueda en cada base de datos Y el período buscado.
- Anote cualquiera de las formas de restricción de lenguaje o publicación (pero debe referirse a la Sección 6.4.9).
- Enumere los individuos o las organizaciones contactados.

Para mayor información sobre cómo debería enumerarse esta información ver el Capítulo 11 (Sección 11.8).

Informar el proceso de búsqueda en la sección de Métodos

En la(s) sección(es) 'Métodos de búsqueda para la identificación de estudios':

- Enumere todas las bases de datos consultadas.
- Anote las fechas de la última búsqueda para cada base de datos Y el período buscado.
- Anote cualquiera de las formas de restricción del lenguaje o de la publicación (pero debe referirse a la Sección 6.4.9).
- Enumere las Fuentes de literatura gris.
- Enumere los individuos o las organizaciones contactados.
- Enumere algunas de las revistas o actas de conferencias que tuvieron búsqueda manual específicamente para esta revisión.
- Enumere algunas otras fuentes consultadas (por ejemplo, listas de referencias, Internet).

Las estrategias de búsqueda completas para cada base de datos deberían incluirse en un Anexo de la revisión para no interrumpir el flujo del texto de la revisión. Las estrategias de búsqueda deberán ser copiadas exactamente como se han llevado a cabo e incluidas en su totalidad junto con los números de línea para cada serie de la búsqueda. No deberían escribirse de nuevo porque esto puede introducir errores. Para una guía más detallada en este tema, contactar al Coordinador de Búsqueda de Ensayos.

Reportar el proceso de búsqueda en la sección Resultados
El número de datos recuperados por las búsquedas electrónicas debe incluirse en la sección Resultados.

Informar la fecha de la búsqueda
Una sola fecha debería especificarse en el campo 'Fecha de búsqueda', para indicar cuando comenzó la búsqueda extensa más reciente. Mayor información para especificar esta fecha está en el Capítulo 3 (Sección 3.3.3).

6.6.3 Puntos de resumen

- Buscar ayuda para documentar el proceso de búsqueda de un Coordinador de Búsqueda de Ensayos antes de comenzar la búsqueda.
- La estrategia completa para cada búsqueda de cada base de datos debería ser copiada en un Anexo de la revisión.
- El número total de datos recuperados por cada estrategia de búsqueda deberá incluirse en la sección de Resultados.
- Guarde enseguida o imprima copias de archivos de cualquier información que encuentre en Internet, como información sobre ensayos en proceso.
- Consultar el Capítulo 4 (Sección 4.5) y el Capítulo 11 (Sección 11.8) para más información sobre qué incluir en la revisión y en el resumen, respectivamente.

6.7 Información del capítulo

Autores: Carol Lefebvre, Eric Manheimer y Julie Glanville en representación del Grupo Cochrane de Métodos de Recuperación de Información.

La versión en inglés de este capítulo debe citarse como: Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: Este capítulo se ha elaborado con secciones de ediciones previas del *Manual* realizado desde 1995 en colaboración con Kay Dickersin, Kristen Larson, Carol Lefebvre y Eric Manheimer. Muchas de las fuentes enumeradas en este capítulo las hemos podido conocer a través de los años gracias a varias personas a quienes queremos agradecer esta ayuda. Queremos agradecerles a los especialistas en información que han compartido con nosotros información y documentación sobre sus procesos de búsqueda. También queremos agradecer a los Coordinadores de Cochrane para la Búsqueda de Ensayos, a los Miembros del Grupo de Métodos de Recuperación de la Información (ver Recuadro 6.7.a), al Grupo Internacional de Interés Especial de la Evaluación de Tecnología en Salud en Fuentes de Información y al Subgrupo de Especialistas en Información InterTASC por sus comentarios sobre los diseños iniciales de este Capítulo, a Anne Eisinga la lectura de las pruebas de las estrategias de búsqueda, y a los dos revisores por pares, Steve McDonald y Ruth Mitchell, por sus comentarios detallados y constructivos.

Recuadro 6.7.a: El Grupo Cochrane de Métodos de Recuperación de Información

El Grupo de Métodos para Recuperar Información (IRMG, sigla en ingles) intenta aportar consejo y apoyo para realizar la búsqueda y facilitar el intercambio de información buscando métodos para apoyar las actividades de recuperación de la información de la Colaboración Cochrane. El Grupo fue registrado oficialmente con la Colaboración Cochrane en Noviembre de 2004. Sus miembros se dedican a aportar apoyo práctico para el desarrollo de técnicas de recuperación de información y facilitar la información a los buscadores. Los objetivos del grupo se logran con las actividades siguientes:

Ofrecer consejo sobre política y práctica de recuperación de la información;

Dar formación y apoyo:

Realizar investigación empírica (incluidas las revisiones sistemáticas) en métodos de recuperación de información;

Ayudar a monitorizar la calidad de las técnicas de búsqueda utilizadas en las revisiones sistemáticas;

Asociarse con miembros de la Colaboración Campbell para evitar la duplicación de esfuerzos en las áreas de recuperación de información que interesen tanto a la Colaboración Cochrane como a la Campbell;

Servir como foro de discusión.

Sitio Web: www.cochrane.org/docs/irmg.htm

6.8 Referencias

Bennett 2003
Bennett DA, Jull A. FDA: untapped source of unpublished trials. *The Lancet* 2003; 361: 1402-1403.

De Angelis 2004
De Angelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, Kotzin S, Laine C, Marusic A, Overbeke AJ, Schroeder TV, Sox HC, Van der Weyden MB, International Committee of Medical Journal Editors. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *JAMA* 2004; 292: 1363-1364.

De Angelis 2005
De Angelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, Kotzin S, Laine C, Marusic A, Overbeke AJ, Schroeder TV, Sox HC, Van der Weyden MB, International Committee of Medical Journal Editors. Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *JAMA* 2004; 293: 2927-2929.

Dickersin 1994
Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; 309: 1286-1291.

Dickersin 2002
Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S, CENTRAL Development Group. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Evaluation and the Health Professions* 2002; 25: 38-64.

Eisinga 2007
Eisinga A, Siegfried N, Clarke M. The sensitivity and precision of search terms in Phases I, II and III of the Cochrane Highly Sensitive Search Strategy for identifying reports of randomized trials in MEDLINE in a specific area of health care - HIV/AIDS prevention and treatment interventions. *Health Information and Libraries Journal* 2007; 24: 103-109.

Eysenbach 2001
Eysenbach G, Tuische J, Diepgen TL. Evaluation of the usefulness of Internet searches to identify unpublished clinical trials for systematic reviews. *Medical Informatics and the Internet in Medicine* 2001; 26: 203-218.

Glanville 2006
Glanville JM, Lefebvre C, Miles JN, Camosso-Stefinovic J. How to identify randomized controlled trials in MEDLINE: ten years on. *Journal of the Medical Library Association* 2006; 94: 130-136.

Glanville 2008
Glanville J, Bayliss S, Booth A, Dundar Y, Fleeman ND, Foster L, Fraser C, Fernandes H, Fry-Smith A, Golder S, Lefebvre C, Miller C, Paisley S, Payne L, Price AM, Welch K, InterTASC Information Specialists' Subgroup. So many filters, so little time: The development of a Search Filter Appraisal Checklist. *Journal of the Medical Library Association* (in press, 2008).

Golder 2006
Golder S, McIntosh HM, Duffy S, Glanville J, Centre for Reviews and Dissemination and UK Cochrane Centre Search Filters Design Group. Developing efficient search strategies to identify reports of adverse effects in MEDLINE and EMBASE. *Health Information and Libraries Journal* 2006; 23: 3-12.

Greenhalgh 2005
Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ* 2005; 331: 1064-1065.

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: 374-380.

Hopewell 2007a
Hopewell S, Clarke M, Lefebvre C, Scherer R. Handsearching versus electronic searching to identify reports of randomized trials. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art No: MR000001.

Hopewell 2007b
Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art No: MR000010.

Horton 1997
Horton R. Medical editors trial amnesty. *The Lancet* 1997; 350: 756.

Khan 2001
Khan KS, ter Riet G, Glanville J, Sowden AJ, Kleijnen J (editors). *Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for those Carrying Out or Commissioning Reviews (CRD Report Number 4)* (2nd edition). York (UK): NHS Centre for Reviews and Dissemination, University of York, 2001.

Lefebvre 2001
Lefebvre C, Clarke M. Identifying randomised trials. In: Egger M, Davey Smith G, Altman DG (editors). *Systematic Reviews in Health Care: Meta-analysis in Context* (2nd edition). London (UK): BMJ Publication Group, 2001.

Lefebvre 2008
Lefebvre C, Eisinga A, McDonald S, Paul N. Enhancing access to reports of clinical trials published world-wide - the contribution of EMBASE records to the Cochrane Central Register of Controlled Trials (CENTRAL) in The Cochrane Library. *Emerging Themes in Epidemiology* (in press, 2008).

MacLean 2003
MacLean CH, Morton SC, Ofman JJ, Roth EA, Shekelle PG. How useful are unpublished data from the Food and Drug Administration in meta-analysis? *Journal of Clinical Epidemiology* 2003; 56: 44-51.

Mallett 2002
Mallett S, Hopewell S, Clarke M. Grey literature in systematic reviews: The first 1000 Cochrane systematic reviews. *Fourth Symposium on Systematic Reviews: Pushing the Boundaries*, Oxford (UK), 2002.

Manheimer 2002
Manheimer E, Anderson D. Survey of public information about ongoing clinical trials funded by industry: evaluation of completeness and accessibility. *BMJ* 2002; 325: 528-531.

McDonald 2002
McDonald S. Improving access to the international coverage of reports of controlled trials in electronic databases: a search of the Australasian Medical Index. *Health Information and Libraries Journal* 2002; 19: 14-20.

Montori 2005
Montori VM, Wilczynski NL, Morgan D, Haynes RB. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ* 2005; 330: 68.

Royle 2003
Royle P, Milne R. Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. *International Journal of Technology Assessment in Health Care* 2003; 19: 591-603.

Sampson 2006
Sampson M, McGowan J. Errors in search strategies were identified by type and frequency. *Journal of Clinical Epidemiology* 2006; 59: 1057-1063.

Scherer 2007
Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art No: MR000005.

Suarez-Almazor 2000
Suarez-Almazor ME, Belseck E, Homik J, Dorgan M, Ramos-Remus C. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. *Controlled Clinical Trials* 2000; 21: 476-487.

White 2001
White VJ, Glanville JM, Lefebvre C, Sheldon TA. A statistical approach to designing search filters to find systematic reviews: objectivity enhances accuracy. *Journal of Information Science* 2001; 27: 357-370.

Whiting 2008
Whiting P, Westwood M, Burke M, Sterne J, Glanville J. Systematic reviews of test accuracy should search a range of databases to identify primary studies. *Journal of Clinical Epidemiology* 2008; 61: 357.e1-357.e10.

Wilczynski 2007
Wilczynski NL, Haynes RB, Hedges Team. EMBASE search strategies achieved high sensitivity and specificity for retrieving methodologically sound systematic reviews. *Journal of Clinical Epidemiology* 2007; 60: 29-33.

Wong 2006
Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. *Journal of the Medical Library Association* 2006; 94: 41-47.

ÍNDICE
CAPÍTULO 7: SELECCIÓN DE LOS ESTUDIOS Y OBTENCIÓN DE LOS DATOS

PUNTOS CLAVE

7.1 INTRODUCCIÓN

7.2 SELECCIÓN DE LOS ESTUDIOS

7.2.1 Estudios (no informes) como la unidad de interés
7.2.2 Identificación de múltiples informes de un mismo estudio
7.2.3 Un proceso típico para seleccionar los estudios
7.2.4 Implementación del proceso de selección
7.2.5 Selección de los estudios “excluidos”
7.2.6 Medición del acuerdo
7.2.6.1 Cálculos de un estadístico kappa simple
Tabla 7.2.a: Datos para el cálculo de uno estadístico kappa simple
Tabla 7.2.b: Datos de ejemplo para el cálculo de un estadístico kappa simple

7.3 QUÉ DATOS OBTENER

7.3.1 ¿Qué son datos?
Tabla 7.3.a: Lista de verificación de los ítems a considerar para la obtención o la extracción de los datos
7.3.2 Métodos y fuentes potenciales de sesgo
7.3.3. Participantes y ámbito
7.3.4 Intervenciones
7.3.4.1 Integridad de las intervenciones
7.3.5 Medidas de desenlace
7.3.5.1 Desenlaces adversos
7.3.6 Resultados
7.3.7 Otra información a obtener

7.4 FUENTES DE LOS DATOS

7.4.1 Informes
7.4.2 Correspondencia con los investigadores
7.4.3 Datos de los pacientes individuales

7.5 FORMULARIOS DE OBTENCIÓN DE DATOS

7.5.1 Justificación para los formularios de obtención de datos
7.5.2 Formularios de obtención de datos electrónicos frente a manuscritos
7.5.3 Diseño de un formulario de obtención de datos
7.5.4 Codificación y explicaciones

7.6 EXTRACCIÓN DE LOS DATOS DE LOS INFORMES

7.6.1 Introducción
7.6.2 ¿Quién debe extraer los datos?
7.6.3 Preparación de la extracción de los datos
7.6.4 Extracción de los datos de múltiples informes del mismo estudio
7.6.5 Fiabilidad y búsqueda de consenso
7.6.6 Resumen

7.7 EXTRACCIÓN DE LOS RESULTADOS DEL ESTUDIO Y CONVERSIÓN AL FORMATO DESEADO

7.7.1 Introducción
7.7.2 Extracción de los datos para los desenlaces dicotómicos
7.7.3 Extracción de los datos para los desenlaces continuos
7.7.3.1 Cambio postintervención versus cambio respecto los valores iniciales
7.7.3.2 Obtención de las desviaciones estándar a partir de los errores estándar y los intervalos de confianza para las medias de los grupos
7.7.3.3 Obtención de las desviaciones estándar a partir de los errores estándar, los intervalos de confianza, los valores de t y los valores de p para las diferencias de medias
7.7.3.4 Transformaciones y datos asimétricos
7.7.3.5 Medianas y rangos intercuartil
7.7.3.6 Rangos
7.7.3.7 Ninguna información sobre la variabilidad
7.7.3.8 Combinación de los grupos
Tabla 7.7.a: Fórmula para combinar los grupos
7.7.4 Extracción de los datos para desenlaces ordinales
7.7.5 Extracción de los datos de recuento
7.7.5.1 Extracción de recuentos como datos dicotómicos
7.7.5.2 Extracción de recuentos como datos continuos
7.7.5.3 Extracción de recuentos como datos de tiempo hasta el evento
7.7.5.4 Extracción de recuentos como datos de tasa
7.7.6 Extracción de datos para desenlaces de tiempo hasta el evento
7.7.7 Extracción de los datos para las estimaciones de los efectos
7.7.7.1 Estimaciones del efecto y metanálisis genérico del inverso de la varianza
7.7.7.2 Obtención de los errores estándar a partir de los intervalos de confianza y los valores de p: medidas absolutas (diferencia)
7.7.7.3 Obtención de los errores estándar a partir de los intervalos de confianza y los valores de p: medidas de proporción

7.8 TRATAMIENTO DE LOS DATOS

7.9 INFORMACIÓN DEL CAPÍTULO

7.10 REFERENCIAS

CAPÍTULO 7: SELECCIÓN DE LOS ESTUDIOS Y OBTENCIÓN DE LOS DATOS

Editores: Julian PT Higgins y Jonathan J Deeks.

Puntos clave

- La evaluación de la elegibilidad de los estudios y la extracción de los datos de los informes de los estudios se debe realizar por al menos dos personas, de forma independiente.
- Las revisiones Cochrane de Intervención tienen estudios, en lugar de informes, como la unidad de interés, por lo que es necesario vincular y reunir los múltiples informes de un mismo estudio.
- Los formularios de extracción de datos son muy importantes. Se deben diseñar cuidadosamente para lograr los objetivos de la revisión, y se deben someter a una prueba piloto para cada nueva revisión (o equipo de revisión).
- Existen indicaciones para ayudar a diseñar y utilizar los formularios de obtención de datos.
- Los datos se pueden encontrar en varios formatos, pero a menudo se pueden convertir en un formato apropiado para el metanálisis.

7.1 Introducción

Los hallazgos de una revisión sistemática dependen de forma importante de las decisiones con respecto a qué estudios se incluyen, y de las decisiones con respecto a qué datos de estos estudios se presentan y analizan. Los métodos utilizados para estas decisiones deben ser transparentes, y se deben seleccionar con el objetivo de disminuir los sesgos y los errores humanos. Aquí se describen los enfoques que se deben utilizar en las revisiones Cochrane para seleccionar los estudios y decidir cuáles de sus datos se presentan.

7.2 Selección de los estudios

7.2.1 Estudios (no informes) como la unidad de interés

Una revisión Cochrane es una revisión de *estudios* que cumplen criterios especificados con anterioridad para su inclusión en la revisión. Como cada estudio puede haberse publicado en varios artículos, resúmenes u otros informes, una búsqueda exhaustiva de los estudios para la revisión puede identificar muchos *informes* de estudios potencialmente relevantes. Por lo tanto, se requieren dos procesos diferentes para determinar qué estudios se pueden incluir en la revisión. Uno es vincular múltiples informes de un mismo estudio; y el otro es utilizar la información disponible de los diferentes informes para determinar qué estudios son elegibles para inclusión. Aunque en algunas ocasiones existe un informe único para cada estudio, nunca se debe suponer que éste es el caso.

7.2.2 Identificación de múltiples informes de un mismo estudio

La publicación duplicada de estudios puede introducir sesgos significativos si los estudios se incluyen inadvertidamente más de una vez en un metanálisis (Tramèr 1997). La publicación duplicada puede adoptar varias formas, que varían desde manuscritos idénticos hasta informes que describen diferentes números de participantes y diferentes resultados (von Elm 2004). Puede ser difícil detectar las publicaciones duplicadas, y pudiera ser necesario que los revisores realicen “trabajo de detective”.

Algunos de los criterios más útiles para comparar los informes son:

- nombres de los autores (la mayoría de los informes duplicados tienen autores en común, aunque no siempre es el caso);
- lugar y ámbito (particularmente si se nombran instituciones como los hospitales);
- detalles específicos de las intervenciones (p.ej. dosis, frecuencia);
- número de participantes y datos basales; y
- fecha y duración del estudio (lo que también puede aclarar si los tamaños de muestra diferentes se deben a diferentes períodos de reclutamiento).

Cuando después de considerar éstos y otros factores se mantiene la incertidumbre, puede ser necesario intercambiar correspondencia con los autores de los informes.

7.2.3 Un proceso típico para seleccionar los estudios

Un proceso típico para seleccionar los estudios para su inclusión en una revisión es el siguiente (el proceso se debe detallar en el protocolo de la revisión).

1. Integrar los resultados de la búsqueda mediante programas informáticos de gestión de referencias bibliográficas, y eliminar los registros duplicados de un mismo informe (ver Capítulo 6, Sección 6.5).
2. Examinar los títulos y resúmenes para eliminar los informes claramente irrelevantes (en este estadio los revisores deberían, en general, incluir más que excluir).
3. Recuperar el texto completo de los informes potencialmente relevantes.
4. Vincular y reunir los informes múltiples de un mismo estudio (ver Sección 7.2.2).

5. Examinar el texto completo de los informes para verificar el grado de cumplimiento que tienen los estudios de los criterios de elegibilidad.
6. Establecer correspondencia con los investigadores, en caso necesario, para aclarar la elegibilidad del estudio (puede ser conveniente solicitar al mismo tiempo información adicional como puede ser resultados que faltan).
7. Tomar una decisión definitiva sobre la inclusión del estudio y proceder a la obtención de los datos.

7.2.4 Implementación del proceso de selección

Las decisiones acerca de qué estudios incluir en una revisión están entre las decisiones más influyentes que se toman en el proceso de revisión. Sin embargo, requieren un juicio de valor. Para ayudar a que estas valoraciones sean reproducibles es conveniente que partes del proceso las realicen más de un revisor. En la práctica, el enfoque exacto puede variar de revisión en revisión, lo que depende en parte de la experiencia y la habilidad de los revisores.

Los revisores deben decidir primero si más de uno de ellos evaluará los títulos y resúmenes de los registros recuperados de la búsqueda (paso 2 en la Sección 7.2.3). La participación de más de dos revisores reduce la posibilidad de que se descarten informes relevantes (Edwards 2002). Es muy importante que la selección final de los estudios la realice más de un revisor (paso 5 en la Sección 7.2.3).

Los expertos en un área particular frecuentemente tienen opiniones establecidas de antemano que pueden sesgar sus evaluaciones sobre la relevancia y validez de los artículos (Cooper 1989, Oxman 1993). Por lo tanto, aunque es importante que al menos un revisor tenga conocimientos sobre el área a revisar, puede ser beneficioso tener un segundo revisor que no sea experto en el contenido. Algunos revisores pueden decidir que las evaluaciones más relevantes las hagan personas cegadas o enmascaradas a la información del artículo, como la revista en la que se publicó, los autores, la institución y la magnitud y dirección de los resultados. Es posible intentarlo editando las copias de los artículos. Sin embargo, esto toma mucho tiempo y es posible que valga la pena debido a los recursos necesarios y a la incertidumbre sobre su beneficio en cuanto a la protección contra el sesgo (Berlin 1997).

Los desacuerdos sobre si se debe incluir un estudio generalmente se pueden resolver mediante discusión. A menudo la causa del desacuerdo es que simplemente uno de los revisores pasó por alto una parte del artículo. Cuando los desacuerdos se deben a diferencias en la interpretación puede ser necesaria la mediación de otra persona. En algunas ocasiones no será posible resolver los desacuerdos sobre si incluir un estudio sin información adicional. En estos casos, los revisores pueden decidir categorizar el estudio en su revisión como en espera de evaluación, hasta que se obtenga información adicional de los autores del estudio.

En resumen, la sección “Métodos” del protocolo y la revisión deben detallar:

- si más de un revisor examina cada título y resumen para excluir informes evidentemente irrelevantes;
- si quienes examinan cada texto completo del informe para determinar la elegibilidad lo haran de forma independiente (lo deben realizar al menos dos personas);
- si las decisiones anteriores las toman expertos en el área de contenido, metodólogos o ambos;
- si las personas que evaluan la relevancia de los estudios conocen los nombres de los autores, las instituciones, la revista de publicación y los resultados cuando aplican los criterios de elegibilidad; y
- cómo se tratan los desacuerdos.

El no cumplimiento de un único criterio de elegibilidad es suficiente para que un estudio se excluya de la revisión. Por lo tanto, en la práctica los criterios de elegibilidad de cada estudio

se deben evaluar en orden de importancia, de manera que la primera respuesta "no" se pueda utilizar como el motivo primario para la exclusión del estudio, y que no sea necesario evaluar los criterios restantes.

En la mayoría de las revisiones será importante realizar pruebas piloto de los criterios de elegibilidad en una muestra de informes (es decir, diez o 12 artículos, incluyendo algunos definitivamente elegibles, definitivamente no elegibles y dudosos). La prueba piloto se puede utilizar para refinar y aclarar los criterios de elegibilidad, adiestrar a las personas que los aplicarán y asegurar que los criterios los pueda aplicar de forma consistente más de una persona.

7.2.5 Selección de los estudios "excluidos"

Una revisión Cochrane incluye una lista de estudios excluidos, que detalla cualquier estudio que un lector pudiera razonablemente esperar que se encoentrara entre los estudios incluidos. Lo anterior incluye todos los estudios que superficialmente pueda parecer que cumplen los criterios de elegibilidad pero que en una inspección adicional no es así, y también los que no cumplen todos los criterios pero son bien conocidos y es probable que algunos lectores los consideren relevantes. Al enumerar dichos estudios como excluidos y proporcionar el motivo primario de la exclusión, los revisores pueden mostrar que estos estudios se tomaron en consideración. La lista de estudios excluidos debe ser tan breve como sea posible. No de deberían enumerar todos los informes que se identificaron mediante la búsqueda exhaustiva. No se deberían enumerar los que evidentemente no cumplen los criterios de inclusión de la revisión tal como aparecen en "Tipos de estudios", "Tipos de participantes" y "Tipos de intervenciones", y en particular no se deberían listar los estudios que evidentemente son no aleatorios, si la revisión incluye solamente ensayos aleatorios.

7.2.6 Medición del acuerdo

Existen medidas formales de acuerdo disponibles para describir el grado en el cual coinciden las evaluaciones realizadas por múltiples revisores (orwin 1994). En la sección 7.2.6.1 se describe cómo se puede calcular un estadístico kappa para medir el acuerdo entre dos revisores que toman decisiones sencillas sobre la inclusión/exclusión. Se considera que los valores de kappa entre 0,40 y 0,59 reflejan un acuerdo aceptable, entre 0,60 y 0,74 un acuerdo adecuado y 0,75 o más reflejan un acuerdo excelente (Orwin 1994).

No se recomienda el cálculo del estadístico kappa como un procedimiento estándar en las revisiones Cochrane, aunque sea útil para mostrar la existencia de problemas, especialmente en los estadios tempranos de la prueba piloto. Es poco probable que la comparación de un valor de kappa con puntos de corte arbitrarios represente la verdadera repercusión de cualquier desacuerdo en la revisión. Por ejemplo, el desacuerdo sobre la elegibilidad de un estudio grande y bien realizado tiene repercusiones más significativas que el desacuerdo sobre un estudio pequeño con riesgos de sesgo. Se deben explorar los motivos de cualquier desacuerdo. Estos pueden revelar la necesidad de revisar los criterios de elegibilidad o los esquemas de codificación para la obtención de los datos, y se debe informar de cualquier cambio que se realice en consecuencia.

7.2.6.1 Cálculos de un estadístico kappa simple

Suponga que K estudios se distribuyen según los números *a* a *i* como en la Tabla 7.2.a. Entonces

$$\text{kappa} = \frac{P_O - P_E}{1 - P_E},$$

donde

$$P_O = \frac{a + e + i}{K}$$

es la proporción de estudios sobre los cuales hubo acuerdo, y

$$P_E = \frac{I_1 \times I_2 + E_1 \times E_2 + U_1 \times U_2}{K^2}$$

es la proporción de estudios sobre los cuales se esperaría que hubiera acuerdo sólo por azar. Como ejemplo, de los datos en la Tabla 7.2.b,

$$P_O = \frac{5+7+3}{25} = 0,6$$

$$P_E = \frac{12x5+10x10+3x10}{25^2} = 0,304$$

y entonces

$$\text{kappa} = \frac{0,6-0,304}{1-0,304} = 0,43.$$

Tabla 7.2.a: Datos para el cálculo de uno estadístico kappa simple

		Revisor 2			
		Incluye	Excluye	Dudoso	Total
Revisor 1	Incluye	*a*	*b*	*c*	I_1
	Excluye	*d*	*e*	*f*	E_1
	Dudoso	*g*	*h*	*i*	U_1
	Total	I_2	E_2	U_2	K

Tabla 7.2.b: Datos de ejemplo para el cálculo de un estadístico kappa simple

		Revisor 2			
		Incluye	Excluye	Dudoso	Total
Revisor 1	Incluye	5	3	4	12
	Excluye	0	7	3	10
	Dudoso	0	0	3	3
	Total	5	10	10	25

7.3 Qué datos obtener

7.3.1 ¿Qué son datos?

Para los fines de este capítulo, “datos” se define como cualquier información acerca (o que se deriva) de un estudio, incluidos los detalles de los métodos, los participantes, el ámbito, el contexto, las intervenciones, los desenlaces, los resultados, las publicaciones y los

investigadores. Los revisores deben planificar por adelantado qué datos se necesitarán para su revisión sistemática y desarrollar una estrategia para obtenerlos. Las secciones siguientes revisan los tipos de información que se deben buscar, y se resumen en la Tabla 7.3.a. La sección 7.4 revisa las fuentes principales de los datos.

Tabla 7.3.a: Lista de verificación de los ítems a considerar para la obtención o la extracción de los datos

Los ítems que no se encuentran en paréntesis se deben obtener normalmente en todas las revisiones, los ítems en paréntesis pueden ser relevantes para algunas revisiones y no para otras.

Fuente
- ID del estudio (creado por el revisor);
- ID del informe (creado por el revisor);
- ID del revisor (creado por el revisor);
- Cita y detalles de contacto;

Elegibilidad
- Confirmar la elegibilidad para la revisión;
- Motivo de exclusión;

Métodos
- Diseño del estudio;
- Duración total del estudio;
- Generación de la secuencia*;
- Ocultación de la secuencia de asignación*;
- Cegamiento*;
- Otras inquietudes acerca del sesgo*;

Participantes
- Número total;
- Ámbito;
- Criterios diagnósticos;
- Edad;
- Sexo;
- País;
- [Comorbilidad];
- [Características sociodemográficas];
- [Grupo étnico];
- [Fecha del estudio];

Intervenciones
- Número total de grupos de intervención;

Para cada grupo de intervención y de comparación de interés:
- Intervención específica;
- Detalles de la intervención (suficientes para la replicación, de ser posible);
- [Integridad de la intervención];

Desenlaces
- Desenlace y tiempo (i) obtenido; (ii) informado*;

Para cada resultado de interés:
- Definición del desenlace (con los criterios diagnósticos si es relevante);
- Unidad de medición (si es relevante);
- Para las escalas: límite superior e inferior, y si es buena la puntuación alta o la baja;

Resultados
- Número de participantes asignados a cada grupo de intervención;

Para cada desenlace de interés:
- Tamaño de la muestra;
- Participantes perdidos al estudio*;
- Datos resumen para cada grupo de intervención (p.ej. tabla de 2x2 para los datos dicotómicos; medias y DE para los datos continuos);
- [Estimación del efecto con el intervalo de confianza; valor de p];
- [Análisis de subgrupos];

Misceláneas
- Fuente de financiación;
- Conclusiones clave de los autores del estudio;
- Comentarios varios de los autores del estudio;
- Referencias a otros estudios relevantes;
- Correspondencia requerida;
- Comentarios varios de los autores de la revisión.

*Se requiere una descripción completa para los ítems estándar en la herramienta “Riesgo de sesgo” (ver Capítulo 8, Sección 8.5).

7.3.2 Métodos y fuentes potenciales de sesgo

Diferentes métodos de investigación pueden influir en los desenlaces de los estudios al introducir diferentes sesgos en los resultados. Se deben obtener las características básicas del diseño del estudio para presentarlas en la tabla “Características de los estudios incluidos”, y detallar si el estudio es aleatorizado, si el estudio tiene un diseño en conglomerados (cluster) o cruzado (crossover), y la duración del estudio. Si la revisión incluye estudios no aleatorizados, se deben describir las características apropiadas del estudio (ver Capítulo 13, sección 13.4)

También se debe obtener información para facilitar las evaluaciones del riesgo de sesgo en cada estudio incluido mediante la herramienta descrita en el Capítulo 8 (Sección 8.5). La herramienta cubre aspectos como la generación de la secuencia, la ocultación de la secuencia de asignación, el cegamiento, los datos incompletos y un informe selectivo de los resultados. Para cada ítem en la herramienta se requiere una descripción de lo que sucedió en el estudio, que puede incluir citas literales de los informes de los estudios. Puede ser más conveniente obtener la información para la evaluación de los datos incompletos y el informe selectivo del desenlace junto con la información total de los resultados. El Capítulo 8 (Sección 8.3.4) analiza algunos aspectos de la obtención de la información para las evaluaciones del riesgo de sesgo.

7.3.3 Participantes y ámbito

Los detalles de los participantes y el ámbito se obtienen principalmente para su presentación en la tabla "Características de los estudios incluidos". Algunos Grupos Cochrane de Revisión han desarrollado estándares con respecto a qué características se deben obtener. Habitualmente, los aspectos que se deben obtener son los que podrían (o se considera que podrían) afectar la existencia o la magnitud del efecto de una intervención y los que podrían ayudar a los usuarios a evaluar su aplicabilidad. Por ejemplo, si los revisores sospechan que hay diferencias importantes en el efecto de la intervención entre diferentes grupos socioeconómicos (algo poco frecuente), se debe obtener esta información. Si se considera que los efectos de la intervención son constantes entre los grupos, y si dicha información no sería útil para ayudar a aplicar los resultados, la misma no se debería obtener.

Las características de los participantes que con frecuencia son útiles para evaluar la aplicabilidad incluyen edad y sexo, y siempre se debe obtener información resumida acerca de las mismas, si no es evidente a partir del contexto. Es probable que se presenten en diferentes formatos (p.ej. las edades como medias o medianas, con las desviaciones estándar o los rangos; el sexo como porcentajes o recuentos; y algunas de ellas para todo el estudio o para cada grupo de intervención por separado). Los revisores deben obtener valores consistentes de, de ser posible, y decidir si es más relevante resumir las características para el estudio en general o por separado, por ejemplo, por grupo de intervención. Otras características que en algunas ocasiones son importantes incluyen grupo étnico, detalles sociodemográficos (p.ej. nivel educacional) y presencia de afecciones comórbidas.

Si los ámbitos de los estudios pudieran influir en los efectos o la aplicabilidad de la intervención, entonces se debe obtener la información sobre los mismos. Habitualmente los ámbitos de los estudios de intervención en atención sanitaria incluyen hospitales de agudos, urgencias, atención primaria, instituciones como residencias de ancianos, oficinas, escuelas y comunidades. Algunas veces los estudios se realizan en regiones geográficas diferentes con variaciones importantes en las características culturales que podrían afectar la administración de una intervención y su desenlace. La fecha del estudio puede estar asociado con diferencias tecnológicas importantes o con tendencias en el tiempo. Si dicha información es importante para la interpretación de la revisión, la misma se debe obtener.

Los criterios diagnósticos que se utilizaron para definir la afección de interés pueden ser una fuente particularmente importante de diversidad entre los estudios y se deben obtener. Por ejemplo, en una revisión de tratamiento farmacológico para la insuficiencia cardíaca congestiva, es importante conocer cómo se estableció la definición y la gravedad de la insuficiencia cardíaca en cada estudio (p.ej. disfunción sistólica o diastólica, disfunción sistólica grave con fracciones de eyección por debajo del 20%). De manera similar, en una revisión de tratamiento antihipertensivo es importante describir los niveles iniciales de tensión arterial de los participantes.

7.3.4 Intervenciones

Se deben obtener los detalles de todas las intervenciones experimentales y de comparación relevantes para la revisión, principalmente para su presentación en la tabla "Características de los estudios incluidos". Nuevamente, son necesarios los detalles de los aspectos que pudieran afectar la existencia o la magnitud del efecto, o que pudieran ayudar a los usuarios a evaluar su aplicabilidad. De ser posible, se debe buscar información suficiente (y presentarla en la revisión) para la replicación de las intervenciones en estudio, incluyendo cualquier intervención conjunta administrada como parte del estudio.

Para muchos ensayos clínicos de muchas intervenciones no complejas como las intervenciones farmacológicas o físicas, las vías de administración (p.ej. administración oral o intravenosa, técnica quirúrgica utilizada), la dosificación (cantidad o intensidad de cada tratamiento, frecuencia de administración), el momento (p.ej. en las 24 horas a partir del diagnóstico) y la duración del tratamiento pueden ser relevantes. Para las intervenciones complejas como las que evalúan los enfoques psicoterapéuticos, conductuales y educacionales o las estrategias de administración de la atención sanitaria, es importante obtener información acerca de los contenidos de las intervenciones, quién las administró y el formato y momento de la administración.

7.3.4.1 Integridad de las intervenciones

El grado en el cual se implementan los procedimientos o los componentes específicos de la intervención puede tener consecuencias importantes para los hallazgos de un estudio. Lo anterior se describirá como integridad de la intervención; los términos relacionados incluyen cumplimiento y fidelidad. La verificación de la integridad de la intervención puede ser particularmente importante en las revisiones de intervenciones preventivas y de intervenciones complejas, que a menudo se implementan en condiciones que presentan numerosos obstáculos a su administración ideal (Dane 1998). La información acerca de la integridad de la intervención puede ayudar a determinar si los resultados poco prometedores se deben a una intervención deficientemente conceptualizada o a una administración incompleta de los componentes descritos. La evaluación de la implementación de la intervención también revela información importante sobre la factibilidad de una intervención en ámbitos reales, y en particular cuán probable es que la intervención se implemente como se planificó. Si es difícil lograr la implementación completa en la práctica, el programa tendrá una baja factibilidad (Dusenbury 2003).

Dane y Schneider (Dane 1998) describen los siguientes cinco aspectos de la integridad de los programas preventivos:

1. El grado en el cual los componentes de la intervención especificados se administraron como se prescribieron (*adherencia*);
2. Número, duración y frecuencia de implementación de los componentes de la intervención (*exposición*);
3. Aspectos cualitativos de la administración de la intervención no relacionados directamente con la implementación del contenido prescrito, como el entusiasmo y el adiestramiento de quienes implementan la intervención, las estimaciones globales de la efectividad de la sesión y la actitud del líder hacia la intervención (*calidad de la administración*);
4. Medidas de la respuesta del participante a la intervención, que puede incluir indicadores como los niveles de participación y entusiasmo (*respuesta del participante*);
5. Medidas de protección contra la difusión del tratamiento, es decir, asegurar que los sujetos de cada grupo experimental reciban solamente las intervenciones planificadas (*diferenciación del programa*).

La integridad de una intervención se puede monitorizar durante un estudio con el uso de las medidas de proceso, y la retroalimentación a partir de tal evaluación puede dar lugar a una

evolución de la propia intervención. Los estudios de evaluación de procesos se caracterizan por un enfoque flexible de la obtención de los datos y el uso de numerosos métodos para generar varios tipos diferentes de datos. Estos estudios incluyen métodos cuantitativos y cualitativos. Las evaluaciones del proceso se pueden publicar por separado de la evaluación del resultado de la intervención. Cuando se considere importante, los revisores deben plantearse abordar si el ensayo tomó en cuenta, o midió, factores clave del proceso y si los estudios que abordaron cuidadosamente la integridad mostraron un impacto mayor. Las evaluaciones de procesos pueden ser una fuente útil de factores que influyen potencialmente en la efectividad de una intervención. Sin embargo, es de señalar que es posible que las medidas de éxito del cegamiento (p.ej. en ensayos farmacológicos controlados con placebo) no sean evaluables (ver Capítulo 8, Sección 8.11.1).

Un ejemplo de revisión Cochrane que evalúa la integridad de la intervención lo proporciona una revisión de abandono del hábito de fumar durante el embarazo (Lumley 2004). Los autores encontraron que la evaluación del proceso de la intervención se realizó solamente en algunos ensayos, y en otros la implementación fue menor que la ideal (incluidos algunos de los ensayos más grandes). La revisión destacó cómo la transferencia de una intervención de un ámbito a otro puede reducir su efectividad si se cambian los elementos o si algunos aspectos de los materiales son inapropiados en una cultura.

7.3.5 Medidas de desenlace

Los revisores deben decidir por adelantado si obtendrán información acerca de todos los desenlaces medidos en un estudio, o solamente de los desenlaces de interés (especificados con anterioridad) para la revisión. Debido a que en la Sección 7.3.6 se recomienda que sólo se deben obtener los datos de desenlaces especificados con anterioridad, también se recomienda que sólo se describan en detalle los desenlaces enumerados en el protocolo. Sin embargo, una lista completa de los nombres de todos los desenlaces medidos permite una evaluación más detallada del riesgo de sesgo debido al informe selectivo de desenlaces (ver Capítulo 8, Sección 8.13).

La información acerca de los desenlaces que es probable que sea importante incluye:

- definición (método diagnóstico, nombre de la escala, definición del umbral, tipo de conducta);
- momento;
- unidad de medición (si es relevante); y
- para las escalas: límites superior e inferior, y si una puntuación alta o baja es favorable.

Puede ser útil obtener detalles de informes citados asociados con la escalas, ya que muchos contienen información adicional acerca de los límites superiores e inferiores, la dirección del beneficio, los promedios típicos y las desviaciones estándar, magnitudes del efecto mínimamente importantes, e información acerca de la validación.

Las consideraciones adicionales para los desenlaces económicos se discuten en el Capítulo 15 (Sección 15.4.2.) y para los desenlaces informados por el paciente en el Capítulo 17.

7.3.5.1 Desenlaces adversos

La obtención de los desenlaces de los efectos adversos puede presentar dificultades particulares, discutidas en detalle en el Capítulo 14. La información incluida dentro de los términos “efectos adversos”, “reacción farmacológica adversa”, “efectos secundarios”, “efectos tóxicos”, “eventos adversos” y “complicaciones” se puede considerar como potencialmente adecuada para la obtención de datos cuando se evalúan los efectos perjudiciales de una intervención. Además, puede estar poco claro si un desenlace se debe clasificar como un resultado adverso (y el mismo desenlace se puede considerar como un

efecto adverso en algunos estudios y no en otros). Que no se mencionen los efectos adversos no quiere decir necesariamente que no hayan ocurrido efectos adversos. Habitualmente es más seguro suponer que no se evaluaron o no se registraron. Las medidas de calidad de vida generalmente son medidas que no examinan específicamente un efecto adverso particular de la intervención. Aunque las escalas de calidad de vida se pueden utilizar para registrar el bienestar general, no se deben considerar sustitutas de una evaluación detallada de la seguridad y la tolerabilidad.

Se deben registrar las definiciones precisas de los desenlaces de los efectos adversos y su intensidad, ya que pueden variar entre los estudios. Por ejemplo, en una revisión de aspirina y hemorragia gastrointestinal, algunos ensayos simplemente informan de sangrados intestinales, mientras que otros informan de categorías específicas de sangrado, como hematemesis, melena y proctorragia (Derry 2000). La definición y el informe de la gravedad de las hemorragias (por ejemplo, importante, grave, que requiere ingreso hospitalario) también varía considerablemente entre los ensayos (Zanchetti 1999). Además, un efecto adverso particular se puede describir o medir de diferentes formas entre los estudios. Por ejemplo, los términos “cansancio”, “fatiga” y “letargia” se podrían utilizar para informar sobre los efectos adversos. Los autores de los estudios también pueden utilizar umbrales de resultados “anormales” (por ejemplo, hipocaliemia diagnosticada con una concentración sérica de potasio de 3,0 mmol/l o 3,5 mmol/l).

7.3.6 Resultados

Los resultados se deben obtener solamente del desenlace de interés especificado en el protocolo. Los resultados de otros desenlaces no se deben obtener a menos que el protocolo se modifique para agregarlos, y se debe informar de esta modificación en la revisión. Sin embargo, los revisores deben estar alerta ante la posibilidad de hallazgos importantes e inesperados, particularmente efectos adversos graves.

A menudo los informes de los estudios incluyen varios resultados para el mismo desenlace. Por ejemplo, se pueden utilizar diferentes escalas de medición, los resultados se pueden presentar por separado para diferentes subgrupos y los desenlaces se pueden haber medido en diferentes puntos temporales. La variación en los resultados puede ser muy amplia, según los datos que se seleccionen (Gøtzsche 2007) y los protocolos deberían ser tan específicos como sea posible acerca de qué medidas de desenlace, el tiempo y estadísticas de resumen (p.ej. valores finales frente a cambio respecto los valores iniciales) se van a obtener. Podria ser necesario refinar el protocolo para facilitar las decisiones sobre qué resultados extraer.

La sección 7.7 describe los números que se necesitarán para realizar el metanálisis. La unidad de análisis (p.ej. participante, grupo, parte del cuerpo, período de tratamiento) se debe registrar para cada resultado si la misma no es evidente (ver Capítulo 9, Sección 9.3). El tipo de dato del desenlace determina la naturaleza de los números que se buscarán para cada desenlace. Por ejemplo, para un desenlace dicotómico (“sí” o “no”), se buscarán el número de participantes y el número que presentó el desenlace para cada grupo. Es importante obtener el tamaño de la muestra relevante para cada resultado, aunque no siempre es evidente. Elaborar un diagrama de flujo, como se recomienda en la Declaración CONSORT (Moher 2001), puede ayudar a determinar el flujo de participantes a lo largo del estudio si no hay uno disponible en un informe publicado (disponible en www.consort-statement.org).

Los números necesarios para el metanálisis no siempre están disponibles, pero algunas veces se puede obtener otro estadístico y convertirlo al formato requerido. Por ejemplo, para un desenlace continuo, habitualmente es más conveniente buscar el número de participantes, la media y la desviación estándar para cada grupo de intervención. Con frecuencia no están directamente disponibles, especialmente la desviación estándar, y estadísticos alternativos permiten el cálculo o la estimación de la desviación estándar faltante (como un error estándar, un intervalo de confianza, una prueba estadística (p.ej. [una prueba t o una prueba

F] o un valor de p). Los detalles se proporcionan en la Sección 7.7. Otras consideraciones para tratar los datos no disponibles se discuten en el Capítulo 16 (Sección 16.1).

7.3.7 Otra información a obtener

Será necesario obtener información adicional de cada informe de estudio, incluidos la citación, los detalles de contacto de los autores del estudio y cualquier otro detalle de las fuentes de información adicional acerca del mismo (por ejemplo, un identificador para el estudio que permita encontrarlo en un registro de ensayos clínicos). En muchas áreas es de particular importancia la fuente de financiamiento del estudio, o los potenciales conflictos de intereses de los autores del estudio. Algunos revisores querrán obtener información sobre las características del estudio que influyan en la calidad de su realización, pero que es poco probable que representen directamente un riesgo de sesgo, por ejemplo, si se obtuvo la aprobación ética y si se realizó el cálculo del tamaño de muestra.

Se recomienda que los revisores obtengan las conclusiones clave del estudio incluido tal cómo las informaron los autores. No es necesario informar de estas conclusiones en la revisión, pero se deben utilizar para verificar los resultados de los análisis realizados por los revisores, particularmente los relacionados con la dirección del efecto. Se deben señalar otros comentarios de lo autores del estudio, por ejemplo, cualquier explicación que proporcionen sobre hallazgos inesperados. Las referencias a otros estudios que se citan en el informe del estudio pueden ser útiles, aunque los revisores deben estar al tanto de la posibilidad de sesgo de citación (ver Capítulo 10, Sección 10.2.2.3)

7.4 Fuentes de los datos

7.4.1 Informes

La mayoría de las revisiones Cochrane obtiene la mayor parte de sus datos de los informes de los estudios. Los informes de los estudios incluyen artículos de revistas, libros, disertaciones, resúmenes de congresos y sitios web. Sin embargo, es de señalar que estos son muy variables en cuanto a su fiabilidad, así como a su nivel de detalle. Por ejemplo, los resúmenes de congresos pueden presentar los hallazgos preliminares y puede ser necesario confirmar los resultados finales. Se recomienda firmemente que se utilice un formulario de obtención de datos para extraer los datos de los informes de los estudios (ver Sección 7.6).

7.4.2 Correspondencia con los investigadores

A menudo los revisores encontrarán que no les es posible extraer toda la información clave que ellos buscan a partir de los informes disponibles con respecto a los detalles del estudio y a los resultados numéricos. En tales circunstancias se recomienda a los revisores que establezcan contacto con los investigadores originales. Los revisores necesitarán considerar si su solicitud de información a los autores del estudio incluirá preguntas abiertas, solicitará piezas específicas de información, incluirá un formulario de obtención de datos (incompleto o parcialmente completo), o buscará datos a nivel de los participantes individuales. Los detalles de contacto de lo autores, si no están disponibles en los informes de los estudios, a menudo se pueden obtener de publicaciones alternativas recientes, de la lista del personal de las universidades o mediante una búsqueda general en la *world wide web*.

7.4.3 Datos de los pacientes individuales

En lugar de extraer los datos de las publicaciones de los estudios, los datos de la investigación original pueden ser solicitados directamente a los investigadores responsables de cada estudio. Las revisiones de datos de pacientes individuales (DPI), en las cuales se proporcionan datos de cada uno de los participantes en cada uno de los estudios, son el valor

de referencia (*gold standard*) en términos de disponibilidad de datos. Los DPI se pueden reanalizar centralmente y, de ser apropiado, combinarse en los metanálisis. Las revisiones de DPI se abordan en el Capítulo 18.

7.5 Formularios de obtención de datos

7.5.1 Justificación para los formularios de obtención de datos

El formulario de obtención de datos es un puente entre lo que informan los investigadores originales (p.ej. en los artículos de revistas, resúmenes, correspondencia personal) y lo que finalmente informan los revisores. El formulario de obtención de datos cumple varias funciones importantes (Meade 1997). Primero, el formulario se deriva directamente de la pregunta de revisión y los criterios para evaluar la elegibilidad de los estudios y proporciona un resumen claro de éstos que se puede aplicar a los informes de los estudios identificados. Segundo, el formulario de obtención de datos es el registro histórico de la gran cantidad de decisiones (y cambios en las decisiones) que ocurren durante el proceso de revisión. Tercero, el formulario es la fuente de los datos para su inclusión en un análisis.

Debido a las importantes funciones de los formularios de obtención de datos, se debe invertir suficiente tiempo y reflexión en su diseño. Como cada revisión es diferente, los formularios de obtención de datos variarán entre las revisiones. Sin embargo, hay muchas similitudes en los tipos de información que son importantes, y los formularios se pueden adaptar de una revisión a la próxima. Aunque el término “formulario de obtención de datos” se utiliza en singular, en la práctica puede haber una serie de formularios utilizados para diferentes fines, por ejemplo, un formulario separado para evaluar la elegibilidad de los estudios para su inclusión en la revisión y facilitar la determinación rápida de los estudios que se deben excluir.

7.5.2 Formularios de obtención de datos electrónicos frente a manuscritos

La decisión entre obtener los datos mediante formularios manuscritos y obtenerlos mediante formularios electrónicos depende en gran medida de la preferencia de los revisores. Las ventajas potenciales de los manuscritos incluyen:

- conveniencia o preferencia;
- la extracción de los datos se puede realizar prácticamente en cualquier lugar;
- son más fáciles de crear e implementar (no necesitan programación informática ni programas informáticos especializados);
- proporcionan un registro permanente de todas las manipulaciones y modificaciones (siempre que dichas manipulaciones y modificaciones no se borren);
- comparación sencilla de los formularios completados por revisores diferentes.
- Las ventajas potenciales de los formularios electrónicos incluyen:
- conveniencia o preferencia;
- combinan la extracción de los datos y la grabación de los datos en un solo paso;
- los formularios se pueden programar (p.ej mediante Microsoft Access) para “guiar” al autor a través del proceso de obtención de datos, por ejemplo, señalando preguntas que dependan de las respuestas a preguntas previas;
- los datos de las revisiones que incluyen gran cantidad de estudios son más fáciles de almacenar, ordenar y recuperar;
- permiten conversiones sencillas en el momento de la extracción de los datos (p.ej. desviaciones estándar a partir de los errores estándar; libras a kilogramos);
- comparación rápida de los formularios completados por diferentes autores; y
- consideraciones medioambientales.

Se han desarrollado sistemas electrónicos que ofrecen la mayoría de las ventajas de ambos enfoques (incluido el programa informático comercial SRS: ver www.trialstat.com). Si los

revisores planifican desarrollar sus propios formularios electrónicos mediante hojas de cálculo o programas de bases de datos, se recomienda que (i) se diseñe primero un formulario manuscrito y se someta a una prueba piloto con más de un autor y varios informes de estudios; (ii) la grabación de los datos se estructure de una forma lógica y se codifiquen las respuestas de una forma tan consistente y directa como sea posible; (iii) se verifique la compatibilidad con RevMan; y (iv) se consideren mecanismos para el registro, la evaluación y la corrección de los errores de la grabación de los datos.

7.5.3 Diseño de un formulario de obtención de datos

Cuando se adapta o diseña un formulario de obtención de datos, los revisores primero deben considerar cuánta información se debe obtener. La obtención de demasiada información puede dar lugar a formularios más largos que el informe del estudio original, y puede suponer una pérdida de tiempo. La obtención de muy poca información o la omisión de datos clave pueden dar lugar a la necesidad de regresar al informe del estudio más tarde durante el proceso de revisión.

A continuación se muestran algunos consejos para el diseño de un formulario de obtención de datos, basados en la acumulación informal de la experiencia de numerosos revisores. También se debe consultar la lista de verificación en la Tabla 7.3.a.

- Incluya el título de la revisión o un identificador único. Los formularios de obtención de datos son adaptables entre las revisiones y algunos revisores participan en múltiples revisiones.
- Incluya una fecha de revisión o un número de versión para el formulario de obtención de datos. En ocasiones es necesario revisar los formularios y esta medida reduce la posibilidad de utilizar por error un formulario desactualizado.
- Registre el nombre (o identificación) de la persona que completó el formulario.
- Deje espacio para notas cerca del inicio del formulario. Lo anterior evita colocar notas, preguntas o recordatorios en la última página del formulario, donde es menos probable que se noten. Las notas importantes se pueden grabar en RevMan en la columna “Notas” de la tabla “Características de los estudios incluidos”, o en el texto de la revisión.
- Incluya un identificador único del estudio, así como un identificador único del informe. Esta medida proporciona un vínculo entre múltiples registros del mismo estudio. A cada estudio incluido se le debe dar el identificador de estudio que se utilice en RevMan (habitualmente incluye el apellido del autor principal y el año de la referencia primaria del estudio).
- Incluya la evaluación (o verificación) de la elegibilidad del estudio para la revisión cerca del inicio del formulario. Las primeras secciones del formulario se pueden utilizar luego para el proceso de evaluación de la elegibilidad. Los motivos de la exclusión de un estudio se pueden deducir fácilmente de estas valoraciones. Por ejemplo, si sólo son elegibles los ensayos clínicos verdaderamente aleatorizados, una pregunta del formulario de obtención de datos puede ser: “¿Aleatorizado? Sí, No, Dudoso. Si un estudio utilizó una asignación alterna, la repuesta a la pregunta es “No”, y esta información se puede grabar en la tabla “Características de los estudios excluidos” como el motivo de la exclusión.
- Registre la fuente de cada pieza clave de información obtenida, incluyendo si la misma se encontró en un informe (lo anterior se puede hacer al destacar los datos en copias impresas, por ejemplo) o si la información se obtuvo de fuentes no publicadas o comunicaciones personales. Cualquier información no publicada que se utilice se debe codificar de la misma manera que la información publicada.
- Utilice casillas de verificación o respuestas codificadas para ahorrar tiempo.
- Incluya las opciones “no informado” o “poco claro” junto con las respuestas “sí” o “no”.
- Considere que el formato de las secciones para obtener los resultados concuerden con las tablas de datos de RevMan. Sin embargo, los formularios de obtención de datos deben tener suficiente flexibilidad para permitir cierta variación en la forma en la que se informan los datos. Se recomienda firmemente que los datos de resultado se obtengan en el formato en el cual se informaron (y luego se transformen en un paso posterior).

- Siempre registre los tamaños de muestra cuando obtenga los datos de resultado, además de obtener los números iniciales (p.ej. los participantes asignados al azar). Puede haber diferentes tamaños de muestra para diferentes medidas de desenlace debido a retiradas o exclusiones.
- Deje suficiente espacio para notas.

7.5.4 Codificación y explicaciones

Es importante proporcionar instrucciones detalladas a todos los revisores que utilizarán el formulario de obtención de datos (Stock 1994). Las mismas se pueden insertar junto o cerca del campo de datos del formulario, directamente en la celda que contiene los datos (p.ej. un comentario en Microsoft Excel) o si son muy largas, en una página separada. El uso de esquemas de codificación es eficiente y facilita una presentación sistemática de las características del estudio en la revisión. La codificación exacta es importante y no debe ser tan complicada que confunda fácilmente a la persona que obtiene los datos, o que dé lugar a clasificaciones erróneas. Se debe verificar que los esquemas de codificación se utilicen de forma consistente entre los diferentes revisores.

7.6 Extracción de los datos de los informes

7.6.1 Introducción

En la mayoría de las revisiones Cochrane la fuente primaria de la información acerca de cada estudio son los informes de los estudios publicados, habitualmente en forma de artículos de revistas. Una de las partes más importantes y que consume más tiempo de una revisión sistemática es la extracción de los datos de dichos informes. Habitualmente cuando se diseña el formulario de obtención de datos se tiene en mente la extracción de los datos.

Las búsquedas electrónicas por texto pueden proporcionar una ayuda valiosa para localizar la información dentro de un informe, por ejemplo, al utilizar las facilidades de búsqueda en los visores PDF, en los buscadores de Internet y en los programas informáticos de procesamiento de texto. Sin embargo, la búsqueda por texto no se debe considerar un sustituto de la lectura del informe, ya que la información se puede presentar mediante terminologías variables.

7.6.2 ¿Quién debe extraer los datos?

Se recomienda firmemente que más de una persona extraiga los datos de cada informe para disminuir los errores y reducir los sesgos potenciales que puedan introducir los revisores. Como mínimo, al menos dos personas deben extraer de forma independiente la información que requiere interpretación subjetiva y la información fundamental para la interpretación de los resultados (p.ej. los datos de resultado). Al igual que en la implementación de proceso de selección (Sección 7.2.4), es preferible que quienes extraigan los datos sean de disciplinas complementarias, por ejemplo, un metodólogo y un especialista en el área de interés. Es importante que cada uno de los involucrados en la extracción de los datos tenga práctica en el uso del formulario y, si el formulario lo diseñó otra persona, que reciba un adiestramiento apropiado.

Las pruebas que apoyan la extracción de los datos por duplicado provienen de varias fuentes indirectas. Un estudio observó que la extracción independiente de los datos por dos revisores dio lugar a menos errores que la extracción de los datos por un solo revisor seguida de la verificación por un segundo revisor (Buscemi 2006). Se ha observado una alta prevalencia de errores en la extracción de los datos (errores en 20 de 34 revisiones) (Jones 2005). Otro estudio de extracción de datos para calcular las diferencias de medias estandarizadas encontró que como mínimo siete de 27 revisiones tenían errores significativos (Gøtzsche 2007).

7.6.3 Preparación de la extracción de los datos

Se debe realizar una prueba piloto con todos los formularios mediante una muestra representativa de los estudios que se van a revisar. Esta prueba puede identificar los datos no disponibles del formulario, o los que es probable que sean superfluos. Es recomendable utilizar estos informes piloto para los borradores de las entradas en las tablas "Características de los estudios incluidos" (Capítulo 11, Sección 11.2) y "Riesgo de sesgo" (Capítulo 8, Sección 8.5). Los usuarios del formulario pueden proporcionar retroalimentación acerca de que ciertas instrucciones de codificación son confusas o están incompletas (p.ej. puede ser que una lista de opciones no cubra todas las situaciones). Es posible que sea necesario el consenso entre los revisores antes de modificar el formulario para evitar cualquier malentendido o desacuerdos tardíos. Puede ser necesario repetir la prueba piloto en un nuevo conjunto de informes si se necesitan cambios importantes después de una primera prueba.

En ocasiones los problemas con la obtención de los datos aparecerán después de completar la prueba piloto y es posible que sea necesario revisar el formulario después de haber comenzado la extracción de los datos. De hecho, es frecuente que un formulario de obtención de datos necesite modificaciones después de habérsele aplicado una prueba piloto. Cuando se le realizan cambios al formulario o a las instrucciones de codificación, puede ser necesario regresar a los informes en los que ya se realizó la extracción de los datos. En algunas situaciones sólo pudiera ser necesario aclarar las instrucciones de codificación sin modificar el formulario actual de obtención de datos.

Algunos han propuesto que el revisor esté cegado a parte de la información del informe, por ejemplo a los autores, antes de la extracción de los datos y la evaluación del riesgo de sesgo (Jadad 1996); ver también Capítulo 9 (Sección 8.3.4). Sin embargo, generalmente el cegamiento de los revisores a los aspectos del informe del estudio no se recomienda para las revisiones Cochrane (Berlin 1997).

7.6.4 Extracción de los datos de múltiples informes del mismo estudio

Frecuentemente los estudios se informan en más de una publicación (Tramèr 1997, von Elm 2004). Sin embargo, la unidad de interés en una revisión Cochrane de intervención es el estudio y no el informe. Por lo tanto, es necesario cotejar la información de múltiples informes. No es apropiado descartar cualquier informe de un estudio incluido, ya que puede contener información de valor no incluida en el informe primario. Los revisores deberán decidir entre dos estrategias:

- Extraer los datos de cada informe por separado, y luego combinar la información de múltiples formularios de obtención de datos.
- Extraer los datos de todos los informes directamente en un formulario único de obtención de datos.

La decisión de qué estrategia utilizar dependerá de la naturaleza de los informes y puede variar entre los estudios y los informes. Por ejemplo, si están disponibles un artículo de revista completo y múltiples resúmenes de congresos, es probable que la mayor parte de la información se obtenga del artículo de revista, y completar un nuevo formulario de obtención de datos para cada resumen de congreso será una pérdida de tiempo. Por el contrario, si hay dos o más artículos de revista detallados, posiblemente relacionados con diferentes períodos del seguimiento, entonces es probable que sea más fácil realizar la extracción de los datos por separado para estos artículos y cotejar posteriormente la información de los formularios de obtención de datos.

Mostrar los diagramas de flujo para los participantes de un estudio, como se recomienda en la Declaración CONSORT (Moher 2001), puede ser particularmente útil cuando se coteja la información de múltiples informes.

7.6.5 Fiabilidad y búsqueda de consenso

Cuando más de un revisor extrae los datos del mismo informe existe la posibilidad de desacuerdo. Se debe identificar en el protocolo un procedimiento explícito o una regla de decisión para identificar y resolver los desacuerdos. Con más frecuencia la fuente de desacuerdo es un error de uno de los que extraen los datos y se resuelve fácilmente. Por lo tanto, un primer paso importante es la discusión entre los revisores. Con menos frecuencia un desacuerdo puede necesitar de la mediación de otra persona. Cualquier desacuerdo que no se pueda resolver se debe abordar mediante el contacto con los autores del estudio; si lo anterior no tiene éxito, el desacuerdo se debe informar en la revisión.

Se debe registrar cuidadosamente la presencia y la solución de los desacuerdos. Mantener una copia de los datos "como se extrajeron" (además de los datos de consenso) permite la evaluación de la fiabilidad de la codificación. Ejemplos de cómo se puede lograr esto incluyen:

- Utilizar el formulario de obtención de datos de un revisor (en papel) y registrar los cambios después del consenso en un color de tinta diferente.
- Utilizar un formulario separado (en papel) para los datos de consenso.
- Grabar los datos de consenso en un formulario electrónico.

Se puede cuantificar el acuerdo de los ítems codificados, por ejemplo, mediante el estadístico kappa (Orwin 1994), aunque es algo que no se hace de forma sistemática en las revisiones Cochrane. En la Sección 7.2.6 se describe un cálculo sencillo para el acuerdo entre dos revisores. Si se evalúa el acuerdo, dicha evaluación se debe realizar solamente para los datos más importantes (p.ej. evaluaciones clave del riesgo de sesgo, o disponibilidad de resultados clave).

Sin embargo, durante todo el proceso de revisión se debe tener en mente la evaluación informal de la fiabilidad de la extracción de los datos. Por ejemplo, si después de alcanzar el consenso en los primeros estudios los revisores notan un desacuerdo frecuente para datos específicos, entonces es posible que sea necesario modificar las instrucciones de codificación. Además, la estrategia de codificación de un autor puede cambiar con el tiempo, ya que se olvidan las reglas de codificación, lo que indica la necesidad de un readiestramiento y, posiblemente, una recodificación.

7.6.6 Resumen

En resumen, la sección "Métodos" del protocolo y la revisión debe detallar:

- las categorías de los datos que se van a obtener;
- cómo se controlará la verificación de los datos extraídos de cada informe (p.ej. extracción por dos revisores, de forma independiente);
- si la extracción de los datos la realizó un experto en el área de contenido, un metodólogo o ambos;
- pruebas piloto, adiestramiento y existencia de instrucciones de codificación para el formulario de extracción de datos;
- cómo se extraen los datos de múltiples informes del mismo estudio; y
- cómo se trataron los desacuerdos si más de un autor extrajo los datos de cada informe.

7.7 Extracción de los resultados del estudio y conversión al formato deseado

7.7.1 Introducción

Ahora se describen los datos que es necesario obtener de cada estudio para los análisis de los desenlaces dicotómicos, los desenlaces continuos y otros tipos de datos de desenlace. Estos

tipos de datos se analizan en el Capítulo 9 (Sección 9.2). Habitualmente es conveniente obtener los datos resumen por separado para cada grupo de intervención y grabarlo en RevMan, donde es posible calcular las estimaciones del efecto. En algunas ocasiones es posible obtener los datos necesarios sólo de forma indirecta, y puede que los resultados relevantes no sean obvios. Esta sección proporciona algunas indicaciones y técnicas útiles para resolver algunas de estas situaciones. De no ser posible obtener los datos resumen de cada grupo de intervención, las estimaciones del efecto se pueden presentar directamente. En la Sección 7.7.7 se describe cómo se pueden obtener los errores estándar de dichas estimaciones del efecto a partir de los intervalos de confianza y los valores de p.

7.7.2 Extracción de los datos para los desenlaces dicotómicos

Los datos dicotómicos se describen en el Capítulo 9, Sección 9.2.2, y su metanálisis se describe en el Capítulo 9, Sección 9.4.4. Los únicos datos necesarios para un desenlace dicotómico son los números en cada una de las dos categorías de desenlaces en cada uno de los grupos de intervención (los números necesarios para completar las cuatro celdas S_E, F_E, S_C, F_C en el Capítulo 9, Cuadro 9.2.a). Los mismos se graban en RevMan como los números de individuos con los desenlaces y los tamaños totales de muestra para los dos grupos. Es más fiable obtener los datos de los desenlaces dicotómicos como los números que presentaron y no presentaron específicamente el desenlace en cada grupo. Aunque en teoría lo anterior equivale a obtener los números totales y los números que presentaron el desenlace, no siempre está claro si los números totales informados son los de los sujetos en los cuales se midió dicho desenlace. En ocasiones es necesario derivar los números de los que presentaron el evento de los porcentajes (aunque no siempre está claro qué denominador utilizar, y los porcentajes redondeados pueden ser compatibles con más de un numerador).

Algunas veces el número de participantes y el número de eventos no están disponibles, pero puede haberse informado una estimación del efecto como un odds ratio o una razón de riesgos, por ejemplo en un resumen de congreso. Tales datos se pueden incluir en los metanálisis si se utiliza el método genérico del inverso de la varianza, solamente si se acompañan de medidas de incertidumbre como el error estándar, el intervalo de confianza del 95% o un valor exacto de p: ver Sección 7.7.7.

7.7.3 Extracción de los datos para los desenlaces continuos

Los datos continuos se describen en el Capítulo 9, Sección 9.2.3, y su metanálisis se discute en el Capítulo 9, Sección 9.4.5. Para realizar un metanálisis de los datos continuos mediante las diferencias de medias o las diferencias de medias estandarizadas los revisores deben buscar:

- el valor medio de las mediciones del desenlace en cada grupo de intervención (M_E y M_C);
- la desviación estándar de las mediciones del desenlace en cada grupo de intervención (DE_E y DE_C);
- el número de participantes en los cuales se midió el desenlace en cada grupo de intervención (N_E y N_C).

Debido a un informe deficiente y variable, puede ser difícil o imposible obtener la información necesaria de los datos resumen presentados. Los estudios varían en cuanto a los estadísticos que utilizan para resumir el promedio (algunas veces utilizan las medianas en lugar de las medias) y la variación (algunas veces utilizan los errores estándar, los intervalos de confianza, los rangos intercuartil y los rangos en lugar de las desviaciones estándar). También varía la escala seleccionada para analizar los datos (p.ej. mediciones postintervención versus cambio respecto los valores iniciales; escala bruta frente a escala logarítmica).

Un error particularmente engañoso es malinterpretar el error estándar como la desviación estándar. Desafortunadamente no siempre está claro qué es lo que se informa y puede ser necesario cierto razonamiento inteligente, y la comparación con otros estudios. Las

desviaciones estándar y los errores estándar se confunden ocasionalmente en los informes de los estudios, y la terminología se utiliza de forma inconsistente.

De ser necesario, siempre se les debe solicitar a los autores la información faltante y aclaraciones acerca de los estadísticos utilizados. Sin embargo, para muchas de las medidas de variación existe una relación aproximada o algebraica con la desviación estándar, por lo que es posible obtener el estadístico requerido si no se publicó en el artículo, como se explica en las Secciones 7.7.3.2 a 7.7.3.7. Se pueden encontrar más detalles y ejemplos en otras publicaciones (Deeks 1997a, Deeks 1997b). En el Capítulo 16 (Sección 16.1.3) se analizan las opciones si no se consiguen las desviaciones estándar después de realizar intentos para obtenerlas.

Algunas veces el número de participantes, las medias y las desviaciones estándar no están disponibles, pero puede haberse informado una estimación del efecto como una diferencia de medias o una diferencia de medias estandarizada, por ejemplo en un resumen de congreso. Tales datos se pueden incluir en los metanálisis mediante el método genérico del inverso de la varianza, si se acompañan de medidas de incertidumbre como el error estándar, el intervalo de confianza del 95% o un valor exacto de p. Para una diferencia de medias se debe obtener un error estándar apropiado a partir de un intervalo de confianza, y utilizar los primeros pasos del proceso descrito en la Sección 7.7.3.3. Para las diferencias de medias estandarizadas, ver la Sección 7.7.7.

7.7.3.1 Cambio postintervención versus cambio respecto los valores iniciales

Un aspecto frecuente de los datos continuos es que la herramienta de medición utilizada para evaluar el resultado de cada participante también se valora al inicio, es decir, antes de administrar la intervención. Lo anterior brinda la posibilidad de utilizar las diferencias en los cambios respecto los valores iniciales (también llamada puntuación de cambio) como el resultado primario. Se aconseja a los revisores no centrarse en el cambio respecto los valores iniciales a menos que se haya utilizado este método de análisis en alguno de los informes de los estudios.

Cuando se aborda el cambio respecto los valores iniciales se crea una medición única para cada participante, la cual se obtiene al restar la medición final de la medición inicial o al restar la medición inicial de la medición final. Luego los análisis se realizan como para cualquier otro tipo de variable de desenlace continua, utilizando los cambios en lugar de las mediciones finales.

Con frecuencia los estudios de una revisión habrán utilizado una mezcla de cambios respecto los valores iniciales y los valores finales. Algunos estudios informarán ambos; otros informarán solamente las puntuaciones de cambio o los valores finales. Como se explica en el Capítulo 9 (Sección 9.4.5.2), los valores finales y las puntuaciones de cambio se pueden combinar algunas veces en el mismo análisis, por lo que no es necesariamente un problema. Los revisores pueden encontrar conveniente extraer los datos sobre los resultados del cambio respecto los valores iniciales asi como los valores finales, si están disponibles las medias y las desviaciones estándar requeridas. Un problema clave asociado con la decisión de qué análisis utilizar es la posibilidad de informe selectivo del desenlace con valores más extremos, y los revisores deben buscar indicaciones de si éste puede ser el caso (ver Capítulo 8, Sección 8.13).

Otro problema de la extracción de la información sobre las medidas de cambio respecto el valor inicial es que frecuentemente las mediciones del valor inicial y las mediciones finales se presentarán para números diferentes de participantes, debido a la ausencia en las visitas programadas y los abandonos del estudio. Pudiera ser difícil identificar el subconjunto de participantes que informan las mediciones iniciales y finales, para los cuales es posible calcular las puntuaciones de cambio.

7.7.3.2 Obtención de las desviaciones estándar a partir de los errores estándar y los intervalos de confianza para las medias de los grupos

Una desviación estándar se puede obtener de un error estándar de una media al multiplicarlo por la raíz cuadrada del tamaño de la muestra:

$$SD = SE \times \sqrt{N}$$

Cuando se realiza esta transformación, los errores estándar deben corresponder a medias calculadas dentro de un grupo de intervención y no a la diferencia de medias entre los grupos de intervención.

Los intervalos de confianza de las medias también se pueden utilizar para calcular las desviaciones estándar. Nuevamente, este cálculo se aplica a los intervalos de confianza para los valores de la media calculados dentro de un grupo de intervención y no para las estimaciones de las diferencias entre las intervenciones (para éstas, ver Sección 7.7.3.3). La mayoría de los intervalos de confianza son intervalos de confianza del 95%. Si el tamaño de la muestra es grande (por ejemplo, mayor de 100 sujetos en cada grupo), el intervalo de confianza del 95% tiene una amplitud de 3,92 errores estándar (3,92 = 2 × 1,96). La desviación estándar para cada grupo se obtiene al dividir la amplitud del intervalo de confianza por 3,92 y luego se multiplica por la raíz cuadrada del tamaño de la muestra:

$$DE = \sqrt{N} \text{ x (límite superior – límite inferior)/3,92}$$

Para los intervalos de confianza del 90%, 3,92 se debe reemplazar por 3,29; y para los intervalos de confianza del 99% se debe reemplazar por 5,15.

Si el tamaño de la muestra es pequeño (por ejemplo menos de 60 en cada grupo) entonces los intervalos de confianza se deben haber calculado mediante un valor de una distribución t. Es necesario reemplazar los números 3,92; 3,29 y 5,15 por valores ligeramente mayores específicos de la distribución t, los cuales se pueden obtener de las tablas de la distribución t con grados de libertad iguales al tamaño de la muestra del grupo menos 1. Los detalles relevantes de la distribución t están disponibles como apéndices en muchos libros de texto de estadística, o en programas de hojas de cálculo habituales. Por ejemplo, el valor de t para un intervalo de confianza del 95% de un tamaño de muestra de 25 se puede obtener al teclear =tinv(1-0,95,25-1) en una celda de una hoja de cálculo de Microsoft Excel (el resultado es 2,0639). El divisor 3,92 en la fórmula anterior se reemplazaría por 2 × 2,0639 = 4,128.

Para tamaños de la muestra moderados (por ejemplo entre 60 y 100 en cada grupo) se puede haber utilizado una distribución t o una distribución normal estándar. Los revisores deben examinar los indicios de cuál se ha utilizado, y utilizar la distribución t en caso de duda.

Como ejemplo, considere los datos que se presentan a continuación:

Grupo	Tamaño de la muestra	Media	IC del 95%
Intervención experimental	25	32,1	(30,0; 34,2)
Intervención control	22	28,3	(26,5; 30,1)

Los intervalos de confianza se deben haber basado en las distribuciones t con 24 y 21 grados de libertad, respectivamente. El divisor para el grupo de intervención experimental es 4,128,

como en la explicación anterior. La desviación estándar para este grupo es $\sqrt{25} \times (34,2 - 30,0)/4,128 = 5,09$. Los cálculos para el grupo control se realizan de forma similar.

Es importante verificar que el intervalo de confianza sea simétrico alrededor de la media (la distancia entre el límite inferior y la media es la misma que la distancia entre la media y el límite superior). Si éste no es el caso, el intervalo de confianza puede haberse calculado a partir de valores transformados (ver Sección 7.7.3.4).

7.7.3.3 Obtención de las desviaciones estándar a partir de los errores estándar, los intervalos de confianza, los valores de t y los valores de p para las diferencias de medias

Las desviaciones estándar se pueden obtener de los errores estándar, los intervalos de confianza, los valores de t o los valores de p que se relacionan con las diferencias entre las medias de dos grupos. La propia diferencia de medias (DM) es necesaria para los cálculos a partir del valor de t o el valor de p. En todos los casos es necesario suponer que las desviaciones estándar de las mediciones de los resultados son las mismas para ambos grupos, y la desviación estándar se utilizaría entonces para ambos grupos de intervención. Se describe primero cómo se puede obtener un valor de t de un valor de p, y luego cómo se puede obtener un error estándar de un valor de t o de un intervalo de confianza, y finalmente cómo se obtiene la desviación estándar de un error estándar. Los revisores pueden seleccionar los pasos apropiados en este proceso según los resultados que estén disponibles. Se pueden utilizar métodos relacionados para derivar las desviaciones estándar de ciertos estadísticos F, ya que utilizar la raíz cuadrada de un valor de F puede producir el mismo valor de t. A menudo hay que tomar precauciones para asegurar que se utiliza un valor de F apropiado, y se recomienda la asesoría de un estadístico experto.

Del valor de p al valor de t

Cuando se citen los valores de p obtenidos de las pruebas t, el valor de t correspondiente se puede obtener de una tabla de la distribución t. Los grados de libertad están dados por $N_E + N_C - 2$, donde N_E y N_C son los tamaños de muestra en los grupos experimental y control. Se ilustrará con un ejemplo. Considere un ensayo de una intervención experimental ($N_E = 25$) versus una intervención control ($N_C = 22$), donde la diferencia de medias fue DM = 3,8. Se menciona que el valor de p para la comparación fue p = 0,008; obtenido mediante una prueba t de dos muestras.

El valor de t que corresponde a un valor de p de 0,008 y 25+22–2 = 45 grados de libertad es t = 2,78. Lo anterior se puede obtener de una tabla de la distribución t con 45 grados de libertad o mediante un ordenador (por ejemplo, al teclear=tinv(0,008, 45) en cualquier celda en una hoja de cálculo de Microsoft Excel).

Existen dificultades cuando se presentan los niveles de significación (como p < 0,05 o incluso p = NS, que habitualmente implica p > 0,05) en lugar de los valores exactos de p. Un enfoque conservador sería tomar el valor de p en el límite superior (p.ej. para p < 0,05 tomar p = 0,05; para p < 0,01 tomar p = 0,01 y para p < 0,001 tomar p = 0,001). Sin embargo, esta solución no es aplicable para los resultados que se informan como p = NS: ver Sección 7.7.3.7.

Del valor de t al error estándar

El valor de t es el cociente entre la diferencia de medias y el error estándar de la diferencia de medias. Por lo tanto, el error estándar de la diferencia de medias se puede obtener al dividir la diferencia de medias (DM) por el valor de t:

$$EE = \frac{DM}{t}$$

En el ejemplo, el error estándar de la diferencia de medias se obtiene al dividir 3,8 por 2,78; lo cual resulta en 1,37.

Del intervalo de confianza al error estándar

Si hay un intervalo de confianza del 95% disponible para la diferencia de medias, entonces el mismo error estándar se puede calcular como:

$$EE=(\text{límite superior - límite inferior}/3{,}92$$

siempre que el ensayo sea grande. Para los intervalos de confianza del 90%, 3,92 se debe reemplazar por 3,29 y para el intervalo de confianza del 99% se debe reemplazar por 5,15. Si el tamaño de la muestra es pequeño, entonces los intervalos de confianza se deben calcular mediante una distribución t. Los números 3,92; 3,29 y 5,15 se deben reemplazar por números mayores específicos para la distribución t y el tamaño de la muestra, y se pueden obtener de las tablas de la distribución t con grados de libertad iguales a $N_E + N_C - 2$, donde N_E y N_C son los tamaños de la muestra en los dos grupos. Los detalles relevantes de la distribución t están disponibles como apéndices en muchos libros de texto de estadística, o en programas de hoja de cálculo habituales. Por ejemplo, el valor de t para un intervalo de confianza del 95% de la comparación de una muestra de tamaño 25 con una muestra de tamaño 22 se puede obtener al teclear =tinv(1-0,95,25+22-2) en una celda de una hoja de cálculo de Microsoft Excel.

Del error estándar a la desviación estándar

La desviación estándar intra-grupo se puede obtener del error estándar de la diferencia de medias mediante la siguiente fórmula:

$$SD = \frac{SE}{\sqrt{\frac{1}{N_E} + \frac{1}{N_C}}}$$

En el ejemplo,

$$DE = \frac{1{,}37}{\sqrt{\frac{1}{25} + \frac{1}{22}}} = 4{,}69 .$$

Note que esta desviación estándar es el promedio de las desviaciones estándar de las ramas experimental y control, y se debe grabar en RevMan dos veces (una vez por cada grupo de intervención).

7.7.3.4 Transformaciones y datos asimétricos

Los estadísticos resumen se pueden haber presentado después de haber aplicado una transformación a los datos brutos. Por ejemplo, pueden estar disponibles las medias y las desviaciones estándar de los valores logarítmicos (o, de manera equivalente, una media geométrica y su intervalo de confianza). Dichos resultados se deben recoger, ya que se pueden incluir en los metanálisis, o – bajo ciertas suposiciones – se pueden transformar nuevamente a la escala bruta.

Por ejemplo, un ensayo informó sobre las respuestas de anticuerpos meningocócicos 12 meses después de la vacunación con la vacuna de la meningitis C y una vacuna control (MacLennan 2000), como medias geométricas de valores 24 y 4,2 con intervalos de confianza de 17 a 34 y 3,9 a 4,6 respectivamente. Estos resúmenes se obtuvieron al hallar las medias y

los intervalos de confianza de los logaritmos naturales de las respuestas de anticuerpos (para la vacuna 3,18 [IC del 95%: 2,83 a 3,53], y para el control 1,44 [1,36 a 1,53]), y tomar sus exponenciales (antilogaritmos). Se puede realizar un metanálisis en la escala de estos logaritmos naturales de las respuestas de anticuerpos. Las desviaciones estándar de los datos transformados a logaritmo se pueden derivar del último par de intervalos de confianza mediante los métodos descritos en la Sección 7.7.3.2. Para una discusión adicional del metanálisis de datos asimétricos, ver Capítulo 9 (Sección 9.4.5.3).

7.7.3.5 Medianas y rangos intercuartil

La mediana es muy similar a la media cuando la distribución de los datos es simétrica, por lo que en ocasiones se puede utilizar directamente en los metanálisis. Sin embargo, las medias y las medianas pueden ser muy diferentes entre sí si los datos son asimétricos y a menudo se informa de las medianas *porque* los datos son asimétricos (ver Capítulo 9, Sección 9.4.5.3).

Los rangos intercuartil describen dónde se encuentra el 50% central de los desenlaces de los participantes. Cuando los tamaños de muestra son grandes y la distribución de los desenlaces es similar a la distribución normal, la amplitud del rango intercuartil será aproximadamente 1,35 desviaciones estándar. En otras situaciones, y especialmente cuando la distribución de los desenlaces es asimétrica, no es posible calcular una desviación estándar a partir de un rango intercuartil. Es de señalar que el uso de los rangos intercuartil en lugar de las desviaciones estándar a menudo se toma como un signo de que la distribución de los desenlaces es asimétrica.

7.7.3.6 Rangos

Los rangos son muy inestables y, al contrario de otras medidas de variación, aumentan cuando se incrementa el tamaño de la muestra. Describen los extremos de los resultados observados en lugar de la variación promedio. Los rangos no se deben utilizar para las estimaciones de las desviaciones estándar. Un enfoque frecuente ha sido utilizar el hecho de que, con datos con distribución normal, el 95% de los valores se encontrará dentro de 2xDE a cada lado de la media. Por lo tanto, se puede calcular que la DE será aproximadamente la cuarta parte del rango típico de los valores de los datos. Este método no es sólido y se recomienda que no se utilice.

7.7.3.7 Ninguna información sobre la variabilidad

Si ninguno de los métodos anteriores permite calcular las desviaciones estándar del informe de un ensayo (y la información no está disponible por parte de los autores del ensayo) entonces, para realizar el metanálisis, el autor puede verse forzado a imputar (“rellenar”) los datos no disponibles o excluir el estudio del metanálisis: ver Capítulo 16 (Sección 16.1.3). También se puede utilizar un enfoque narrativo para la síntesis. Es de utilidad tabular los resultados disponibles de todos los estudios incluidos en la revisión sistemática, incluso si no se pueden incluir en un metanálisis formal.

7.7.3.8 Combinación de los grupos

En ocasiones es conveniente combinar dos subgrupos informados en un único grupo. Lo anterior pudiera ser el caso, por ejemplo, si un estudio presenta los tamaños de muestra, las medias y las desviaciones estándar por separado para hombres y mujeres en cada uno de los grupos de intervención. La fórmula en la Tabla 7.7.a se puede utilizar para combinar los números en un único tamaño de muestra, media y desviaciones estándar para cada grupo de

intervención (es decir, combinar los hombres y las mujeres en este ejemplo). Es de señalar que la fórmula de apariencia compleja para la DE genera DE de las mediciones de resultado *como si el grupo combinado nunca se hubiera dividido en dos*. Una aproximación a esta

desviación estándar se obtiene al utilizar la desviación estándar habitual agrupada, lo cual proporciona una ligera subestimación de la desviación estándar esperada.

Estas fórmulas también son apropiadas para utilizarlas en estudios que comparan más de dos intervenciones, con el fin de combinar dos grupos de intervención en un grupo de intervención único (ver Capítulo 16, Sección 16.5). Por ejemplo, "Grupo 1" y "Grupo 2" se podrían referir a dos variantes alternativas de una intervención a la cual los participantes se asignaron al azar.

Si hay más de dos grupos para combinar, la estrategia más sencilla es aplicar la fórmula anterior de forma secuencial (es decir, combinar los grupos 1 y 2 para crear el grupo "1+2", luego combinar el grupo "1+2" y el grupo 3 para crear el grupo "1+2+3" y así sucesivamente).

Tabla 7.7.a: Fórmula para combinar los grupos

	Grupo 1 (p.ej. hombres)	Grupo 2 (p.ej. mujeres)	Grupos combinados
Tamaño de muestra	N_1	N_2	$N_1 + N_2$
Media	M_1	M_2	$\frac{N_1M_1 + N_2M_2}{N_1 + N_2}$
DE	DE_1	DE_2	$\sqrt{\frac{(N_1 - 1)SD_1^2 + (N_2 - 1)SD_2^2 + \frac{N_1N_2}{N_1 + N_2}(M_1^2 + M_2^2 - 2M_1M_2)}{N_1 + N_2 - 1}}$

7.7.4 Extracción de los datos para desenlaces ordinales

Los datos ordinales, cuando los desenlaces se clasifican en varias categorías ordenadas, se describen en el Capítulo 9, Sección 9.2.4, y su metanálisis se discute en el Capítulo 9, Sección 9.4.7. Los datos que es necesario extraer para los desenlaces ordinales dependen de si la escala ordinal se va a dicotomizar para el análisis (ver Sección 7.7.2), se va a tratar como un desenlace continuo (ver Sección 7.7.3) o se va a analizar directamente como datos ordinales. Esta decisión, a su vez, estará influenciada por la forma en la cual los autores de los estudios analizaron sus datos. Por lo tanto, podría ser imposible especificar con anterioridad si la extracción de los datos incluirá el cálculo del número de participantes por encima y por debajo del umbral definido, o de los valores de la media y las desviaciones estándar. En la práctica es razonable extraer los datos en todas las formas en que se proporcionan, ya que no estará claro cuál es la más frecuente hasta que se hayan revisado todos los estudios, y en algunas circunstancias puede ser justificable incluir en la revisión más de una forma de análisis.

Cuando los datos ordinales se dicotomizan y existen varias opciones para seleccionar el punto de corte (o la selección del punto de corte es arbitraria), es posible planificar desde el inicio la exploración en un análisis de sensibilidad del impacto que supone la selección del punto de corte (ver Capítulo 9, Sección 9.7). Para hacerlo es necesario obtener los datos que se utilizarán para cada dicotomización alternativa. Por lo tanto, es preferible registrar los números en cada categoría de las escalas ordinales cortas para evitar tener que extraer los datos de un artículo más de una vez. Este enfoque de registrar todas las clasificaciones también es posible cuando los estudios utilizan escalas ordinales cortas algo diferentes, y no está claro si habrá un punto de corte común entre todos los estudios que se pueda utilizar para la dicotomización.

También es necesario registrar los números en cada categoría de la escala ordinal para cada grupo de intervención si se utilizará el método del odds ratio proporcional (ver Capítulo 9, Sección 9.2.4).

7.7.5 Extracción de los datos de recuento

Los recuentos se describen en el Capítulo 9, Sección 9.2.5, y su metanálisis se discute en el Capítulo 9, Sección 9.4.8. Los datos que son inherentemente recuentos se pueden analizar de varias formas. La decisión fundamental es si el resultado de interés se transforma en dicotómico, continuo, tiempo hasta el evento o tasa. Un error frecuente es tratar los recuentos directamente como datos dicotómicos, y utilizar como tamaños de la muestra el número total de participantes o el número total de, por ejemplo, personas-año de seguimiento. Ninguno de estos enfoques es apropiado para un evento que puede ocurrir más de una vez para cada participante. Lo anterior se hace evidente cuando el número total de eventos excede el tamaño de la muestra, lo que da lugar a resultados no esenciales. Aunque es preferible decidir por adelantado cómo se van a analizar los datos de recuentos, a menudo la decisión está determinada por el formato de los datos disponibles, por lo que no es posible tomar la decisión hasta que se haya revisado la mayoría de los estudios. Por lo tanto, en general los revisores deben extraer los datos de recuentos en la forma en que se informan.

Algunas veces los datos detallados sobre los eventos o las personas-año en riesgo no están disponibles, pero los resultados calculados a partir de ellos sí lo están. Por ejemplo, una estimación de la proporción de la tasa o la diferencia de tasas se puede haber presentado en un resumen de congreso. Dichos datos se pueden incluir en los metanálisis sólo si se acompañan de medidas de incertidumbre como el intervalo de confianza del 95% (ver Sección 7.7.7). A partir de éste es posible obtener un error estándar y utilizar el método genérico del inverso de la varianza para el metanálisis.

7.7.5.1 Extracción de recuentos como datos dicotómicos

Para considerar el resultado como dicotómico el revisor debe determinar el número de participantes en cada grupo de intervención, y el número de participantes en cada grupo de intervención que presentó *al menos un evento* (o algún otro criterio apropiado que clasifique a todos los participantes en uno de los dos grupos posibles). Con este enfoque se pierde cualquier elemento temporal de los datos, aunque es posible crear una serie de desenlaces dicotómicos, por ejemplo, al menos un accidente cerebrovascular durante el primer año de seguimiento, al menos un accidente cerebrovascular durante los dos primeros años de seguimiento, y así sucesivamente. Puede ser difícil obtener tales datos de los informes publicados.

7.7.5.2 Extracción de recuentos como datos continuos

Para extraer recuentos como datos continuos (es decir, número promedio de eventos por paciente), se debe seguir la guía en la Sección 7.7.3, aunque se le debe prestar atención particular a la probabilidad de que los datos sean muy asimétricos.

7.7.5.3 Extracción de recuentos como datos de tiempo hasta el evento

Para eventos poco frecuentes que pueden ocurrir más de una vez, un revisor puede tener que lidiar con estudios que tratan los datos como tiempo hasta el *primer* evento. Para extraer los recuentos como datos de tiempo hasta el evento se debe seguir la guía en la Sección 7.7.6.

7.7.5.4 Extracción de recuentos como datos de tasa

Si es posible extraer el número total de eventos en cada grupo y la cantidad total de personas-año con riesgo en cada grupo, entonces los datos de recuento se pueden analizar

como tasas (ver Capítulo 9, Sección 9.4.8). Cabe destacar que no es necesario el número total de participantes para un análisis de datos de tasa, pero se debe registrar como parte de la descripción del estudio.

7.7.6 Extracción de datos para desenlaces de tiempo hasta el evento

Los desenlaces de tiempo hasta el evento se describen en el Capítulo 9, Sección 9.2.6, y su metanálisis se discute en el Capítulo 9, Sección 9.4.9. El metanálisis de los datos de tiempo hasta el evento frecuentemente incluye la obtención de datos de pacientes individuales de los investigadores originales, el reanálisis de los datos para obtener estimaciones del logaritmo del cociente de riesgos instantáneos (*hazard ratio*) y su error estándar, y posteriormente realizar un metanálisis (ver Capítulo 18). A menudo, utilizar la información resumen de los artículos publicados o los informes de los ensayos para realizar un metanálisis es problemático, ya que habitualmente no se presentan la mayoría de los estadísticos de resúmen apropiados. Se pueden utilizar dos enfoques para obtener las estimaciones de los logaritmos de los cocientes de riesgo instantáneos y sus errores estándar con el fin de incluirlos en un metanálisis mediante el método genérico del inverso de la varianza, independientemente de si se utilizan datos de pacientes individuales o datos agregados. Para una guía práctica, los revisores deben consultar Tierney y cols. (Tierney 2007).

En el primer enfoque se puede obtener una estimación del logaritmo del cociente de riesgos instantáneos a partir de los estadísticos calculadas durante un análisis de rango logarítmico (*log-rank*). Se recomienda la colaboración con un estadístico experto si se sigue este enfoque. El logaritmo del cociente de riesgos instantáneos (experimental con respecto a control) se calcula como $(O - E)/V$, el cual tiene un error estándar $1/\sqrt{V}$, donde O es el número de eventos observados en la intervención experimental, E es el rango logarítmico del número de eventos esperado en la intervención experimental, $O - E$ es el estadístice rango logarítmico y V es la varianza del estadístico rango logarítmico. Por lo tanto, es necesario obtener los valores de $O - E$ y V para cada estudio.

Estos estadísticos se calculan fácilmente si los datos de los pacientes individuales están disponibles, y algunas veces se pueden extraer de los estadísticos publicados y de las curvas de supervivencia (Parmar 1998, Williamson 2002). Alternativamente, en ocasiones se pueden utilizar los datos agregados para cada grupo de intervención en cada ensayo. Por ejemplo, asuma que los datos incluyen el número de participantes que presentaron el evento durante el primer año, el segundo año, etc., así como el número de participantes que no presentaron el evento y aún se siguen al final de cada año. Con estos datos se puede realizar un análisis de rango logarítmico para proporcionar los valores $O - E$ y V, aunque se debe tener precaución al tratar los tiempos censurados. Debido a que el agrupamiento es crudo, la estimación del logaritmo del cociente de riesgos instantáneos es sólo aproximada, y en algunas revisiones se le ha llamado logaritmo del odds ratio (Early Breast Cancer Trialist' Collaborative Group 1990). Si los intervalos de tiempo son largos, un enfoque más apropiado es el que se basa en el intervalo de supervivencia censurado (Collett 1994).

El segundo enfoque se puede utilizar si los autores de los ensayos analizaron los datos mediante el modelo de riesgos proporcionales de Cox, o si un modelo de Cox se ajusta a los datos de pacientes individuales. Los modelos de Cox producen estimaciones directas del logaritmo del cociente de riesgos instantáneos y su error estándar (por lo que se puede realizar un metanálisis de la varianza inversa genérica). Si el cociente de riesgos instantáneos se cita en un informe junto con un intervalo de confianza o un valor de p, las estimaciones del error estándar se pueden obtener como se describe en la Sección 7.7.7.

7.7.7 Extracción de los datos para las estimaciones de los efectos

7.7.7.1 Estimaciones del efecto y metanálisis genérico del inverso de la varianza

En algunas revisiones se obtendrá una estimación general del efecto de cada estudio, en lugar de datos resumen para cada grupo de intervención. Este puede ser el caso, por ejemplo, para los estudios no aleatorizados, los ensayos cruzados, los ensayos aleatorios por conglomerados o los estudios con desenlaces de tiempo hasta el evento. El metanálisis se puede aplicar a dichas estimaciones del efecto si están disponibles los errores estándar, mediante el tipo de desenlace genérico del inverso de la varianza de RevMan (ver Capítulo 9, Sección 9.4.3). Cuando se extraen los datos de estudios no aleatorizados y de algunos estudios aleatorizados, pueden estar disponibles las estimaciones de los efectos ajustados (p.ej. los odds ratios ajustados de los análisis de regresión logística o los odds ratio ajustados de los análisis de regresión de Poisson). El proceso de extracción de los datos y el análisis mediante el método genérico del inverso de la varianza es el mismo que para las estimaciones no ajustadas, aunque se deben registrar las variables que se ajustaron (ver Capítulo 13, Sección 13.6.2).

En ocasiones se pueden buscar los datos resumen para cada grupo de intervención (por ejemplo, el número de eventos y participantes, o las medias y las desviaciones estándar), pero no se pueden extraer. En tales situaciones aún puede ser posible incluir el estudio en un metanálisis mediante el método genérico del inverso de la varianza. Una limitación de este enfoque es que las estimaciones y los errores estándar de la misma medida de efecto se deben calcular para todos los otros estudios en el mismo metanálisis, incluso si los mismos proporcionan los datos resumen por grupo de intervención. Por ejemplo, si los números en cada categoría de desenlace por grupo de intervención se conocen para algunos estudios, pero sólo están disponibles los odds ratio (OR) para otros estudios, entonces será necesario calcular los OR para el primer conjunto de estudios y entrarlos en RevMan como el tipo de desenlace genérico del inverso de la varianza, para permitir un metanálisis con el segundo conjunto de estudios. Se puede utilizar RevMan para calcular estos OR (si se graban como datos dicotómicos) y los intervalos de confianza que presenta RevMan se pueden transformar en errores estándar mediante los métodos siguientes.

Las estimaciones de una medida de efecto de interés se pueden presentar junto con un intervalo de confianza o un valor de p. Habitualmente es conveniente obtener el error estándar de estos números, de manera que se pueda utilizar el tipo de desenlace genérico del inverso de la varianza en RevMan para realizar el metanálisis. El procedimiento para obtener un error estándar depende de si la medida de efecto es una medida absoluta (p.ej. diferencia de medias, diferencia de medias estandarizada, diferencia de riesgos) o una medida de proporción (p.ej. odds ratio, cociente de riesgos, cociente de riesgos instantáneos, cociente de tasas). Estos procedimientos se describen en la Sección 7.7.7.2 y la Sección 7.7.7.3, respectivamente. Sin embargo, para las medidas de desenlace continuas, los casos especiales de la extracción de resultados para una media de un brazo de intervención y para la diferencia entre dos medias se abordan en la Sección 7.7.3.

7.7.7.2 Obtención de los errores estándar a partir de los intervalos de confianza y los valores de p: medidas absolutas (diferencia)

Si el intervalo de confianza del 95% está disponible para una medida absoluta del efecto de la intervención (p.ej. DME, diferencia de riesgos, diferencia de tasas), entonces el error estándar se puede calcular como

EE = (límite superior – límite inferior)/3,92.

Para los intervalos de confianza del 90% se divide por 3,29 en lugar de 3,92; para los intervalos de confianza del 99% se divide por 5,15.

Cuando los valores exactos de p se citan junto con las estimaciones del efecto de la intervención, es posible realizar la estimación de los errores estándar. Aunque todas las pruebas de significación producen valores de p, diferentes pruebas utilizan diferentes enfoques matemáticos para obtener un valor de p. El método mostrado aquí supone que los

valores de p se obtuvieron a través de un enfoque particularmente simple al dividir la estimación del efecto por su error estándar y comparar el resultado (llamado Z) con una distribución normal estándar (a menudo los estadísticos se refieren a esta prueba como la prueba de Wald). Cuando las pruebas de significación utilizan otros enfoques matemáticos, es posible que la estimación de los errores estándar no coincida exactamente con los verdaderos errores estándar.

El primer paso es obtener el valor de Z correspondiente al valor de p informado a partir de una tabla de la distribución normal estándar. El error estándar se puede calcular como

EE = estimación del efecto de la intervención/Z

Como ejemplo, asuma que un resumen de congreso presenta una estimación de una diferencia de riesgos de 0,03 (p = 0,008). El valor de Z que corresponde a un valor de p de 0,008 es Z = 2,652. El mismo se puede obtener de una tabla de la distribución normal estándar o de un ordenador (por ejemplo, al teclear =abs(normsinv(0,008/2) en cualquier celda de una hoja de cálculo de Microsoft Excel). El error estándar de la diferencia de riesgos se obtiene al dividir la diferencia de riesgos (0,03) por el valor de Z (2,652), lo cual resulta en 0,011.

7.7.7.3 Obtención de los errores estándar a partir de los intervalos de confianza y los valores de p: medidas de proporción

El proceso para obtener los errores estándar para las medidas de proporción es similar al de las medidas absolutas, pero con un primer paso adicional. Los análisis de las medidas de proporción se realizan en la escala logarítmica natural (ver Capítulo 9, Sección 9.2.7). Para una medida de proporción como el cociente de riesgos, el odds ratio o el cociente de riesgos instantáneos (los cuales se denotarán aquí genéricamente como RR), primero se calcula

límite inferior = ln(límite de confianza inferior proporcionado para el RR)
límite superior = ln(límite de confianza superior proporcionado para el RR)
estimación del efecto de la intervención = lnRR

Luego se puede utilizar la fórmula de la Sección 7.7.7.2. Es de señalar que los errores estándar se refieren al logaritmo de la medida de proporción. Cuando se utiliza el método genérico del inverso de la varianza en RevMan, los datos se deben grabar en la escala logarítmica natural, es decir, como lnRR y el error estándar del lnRR, como se calcula aquí (ver Capítulo 9, Sección 9.4.3).

7.8 Tratamiento de los datos

Es posible obtener los datos en formularios de obtención de datos en papel y grabarlos directamente en RevMan. Sin embargo, a menudo será necesario o conveniente tratar los datos en un programa estadístico antes de entrarlos en RevMan. Existen varios programas informáticos y de procesamiento de datos que pueden ser útiles para estos procedimientos, incluidos los programas de hojas de cálculo (p.ej. Microsoft Excel) y los programas de bases de datos (p.ej. Microsoft Access). Por ejemplo, la tabulación de la información extraída de los estudios en una hoja de cálculo puede facilitar la clasificación de los estudios en comparaciones y subgrupos. Además, las conversiones estadísticas, por ejemplo de errores estándar a desviaciones estándar, se deberían hacer de forma ideal con un ordenador en lugar de utilizar una calculadora de mano, ya que permite guardar un registro permanente de los números originales y calculados, al igual que de los cálculos reales utilizados.

7.9 Información del capítulo

Editores: Julian PT Higgins y Jonathan J Deeks.

La versión en inglés de este capítulo se debe citar como: Higgins JPT, Deeks JJ (editors). Chapter 7: Selecting studies and collecting data. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: Esta sección se redactó a partir de las versiones anteriores del *Manual*. Para detalles adicionales de los autores y editores anteriores del *Manual*, ver Capítulo 1 (Sección 1.4). Andrew Herxheimer, Nicki Jackson, Yoon Loke, Deirdre Price y Helen Thomas contribuyeron al texto. Stephanie Taylor y Sonja Hood contribuyeron con sugerencias para el diseño de los formularios de obtención de datos. Se agradecen a Judith Anzures, Mike Clarke, Miranda Cumpston y Peter Gøtzsche sus útiles comentarios.

7.10 Referencias

Berlin 1997
Berlin JA. Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. *The Lancet* 1997; 350: 185-186.

Buscemi 2006
Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. *Journal of Clinical Epidemiology* 2006; 59: 697-703.

Collett 1994
Collett D. *Modelling Survival Data in Medical Research*. London (UK): Chapman & Hall, 1994.

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

Dane 1998
Dane AV, Schneider BH. Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review* 1998; 18: 23-45.

Deeks 1997a
Deeks J. Are you sure that's a standard deviation? (part 1). *Cochrane News* 1997; Issue No. 10: 11-12. (Available from www.cochrane.org/newslett/ccnewsbi.htm).

Deeks 1997b
Deeks J. Are you sure that's a standard deviation? (part 2). *Cochrane News* 1997; Issue No. 11: 11-12. (Available from www.cochrane.org/newslett/ccnewsbi.htm).

Derry 2000
Derry S, Loke YK. Risk of gastrointestinal haemorrhage with long term use of aspirin: meta-analysis. *BMJ* 2000; 321: 1183-1187.

Dusenbury 2003
Dusenbury L, Brannigan R, Falco M, Hansen WB. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research* 2003; 18: 237-256.

Early Breast Cancer Trialists' Collaborative Group 1990
Early Breast Cancer Trialists' Collaborative Group. *Treatment of Early Breast Cancer. Volume 1: Worldwide Evidence 1985-1990*. Oxford (UK): Oxford University Press, 1990. (Available from www.ctsu.ox.ac.uk).

Edwards 2002
Edwards P, Clarke M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. *Statistics in Medicine* 2002; 21: 1635-1640.

Gøtzsche 2007
Gøtzsche PC, Hróbjartsson A, Maric K, Tendal B. Data extraction errors in meta-analyses that use standardized mean differences. *JAMA* 2007; 298: 430-437.

Jadad 1996
Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, McQuay H. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials* 1996; 17: 1-12.

Jones 2005
Jones AP, Remmington T, Williamson PR, Ashby D, Smyth RL. High prevalence but low impact of data extraction and reporting errors were found in Cochrane systematic reviews. *Journal of Clinical Epidemiology* 2005; 58: 741-742.

Lumley 2004
Lumley J, Oliver SS, Chamberlain C, Oakley L. Interventions for promoting smoking cessation during pregnancy. *Cochrane Database of Systematic Reviews* 2004, Issue 4. Art No: CD001055.

MacLennan 2000
MacLennan JM, Shackley F, Heath PT, Deeks JJ, Flamank C, Herbert M, Griffiths H, Hatzmann E, Goilav C, Moxon ER. Safety, immunogenicity, and induction of immunologic memory by a serogroup C meningococcal conjugate vaccine in infants: A randomized controlled trial. *JAMA* 2000; 283: 2795-2801.

Meade 1997
Meade MO, Richardson WS. Selecting and appraising studies for a systematic review. *Annals of Internal Medicine* 1997; 127: 531-537.

Moher 2001
Moher D, Schulz KF, Altman DG. The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *The Lancet* 2001; 357: 1191-1194. (Available from www.consort-statement.org).

Orwin 1994
Orwin RG. Evaluating coding decisions. In: Cooper H, Hedges LV (editors). *The Handbook of Research Synthesis*. New York (NY): Russell Sage Foundation, 1994.

Oxman 1993
Oxman AD, Guyatt GH. The science of reviewing research. *Annals of the New York Academy of Sciences* 1993; 703: 125-133.

Parmar 1998
Parmar MKB, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine* 1998; 17: 2815-2834.

Stock 1994
Stock WA. Systematic coding for research synthesis. In: Cooper H, Hedges LV (editors). *The Handbook of Research Synthesis*. New York (NY): Russell Sage Foundation, 1994.

Tierney 2007
Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007: 16.

Tramèr 1997
Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997; 315: 635-640.

von Elm 2004
von Elm E, Poglia G, Walder B, Tramèr MR. Different patterns of duplicate publication: an analysis of articles used in systematic reviews. *JAMA* 2004; 291: 974-980.

Williamson 2002
Williamson PR, Smith CT, Hutton JL, Marson AG. Aggregate data meta-analysis with time-to-event outcomes. *Statistics in Medicine* 2002; 21: 3337-3351.

Zanchetti 1999
Zanchetti A, Hansson L. Risk of major gastrointestinal bleeding with aspirin (Authors' reply). *The Lancet* 1999; 353: 149-150.

ÍNDICE
CAPÍTULO 8:
EVALUACIÓN DEL RIESGO DE SESGO EN LOS ESTUDIOS INCLUIDOS

PUNTOS CLAVE

8.1 INTRODUCCIÓN

8.2 ¿QUÉ ES SESGO?

8.2.1 “Sesgo” y “riesgo de sesgo”
8.2.2 “Riesgo de sesgo” y “calidad”
8.2.3 Establecer la evidencia empírica de la presencia de sesgos

8.3 HERRAMIENTAS PARA EVALUAR LA CALIDAD Y EL RIESGO DE SESGO

8.3.1 Tipos de herramientas
8.3.2 Publicación frente a realización
8.3.3 Escalas de calidad y revisiones Cochrane
8.3.4 Obtención de información para las evaluaciones del riesgo de sesgo

8.4 INTRODUCCIÓN A LAS FUENTES DE SESGO EN LOS ENSAYOS CLÍNICOS

8.4.1 Sesgo de selección
8.4.2 Sesgo de realización
8.4.3 Sesgo de detección
8.4.4 Sesgo de desgaste
8.4.5 Sesgo de notificación
8.4.6 Otros sesgos
Tabla 8.4.a: Un esquema de clasificación frecuente para el sesgo

8.5 HERRAMIENTA DE LA COLABORACIÓN COCHRANE PARA EVALUAR EL RIESGO DE SESGO

8.5.1 Generalidades
Tabla 8.5.a: Herramienta de la Colaboración Cochrane para evaluar el riego de sesgo
Tabla 8.5.b: Diferencias entre la herramienta ‘Riesgo de sesgo’ descrita en el Manual versiones 5.0.1/5.0.2 y la herramienta ‘Riesgo de sesgo’ revisada descrita en el Manual versión 5.1.0. (esta versión)
8.5.2 Ayuda para la evaluación
Tabla 8.5.c: Ejemplos de ayudas para la evaluación de la generación de la secuencia
8.5.3 La evaluación
Tabla 8.5.d: Criterios para evaluar el riesgo de sesgo en la herramienta de evaluación “Riesgo de sesgo”

8.6 PRESENTACIÓN DE LAS EVALUACIONES DEL RIESGO DE SESGO

Tabla 8.6.a: Ejemplo de una tabla de “Riesgo de sesgo” para un estudio único (ficticio)
Figura 8.6.b: Ejemplo de una figura “Gráfico del riesgo de sesgo”
Figura 8.6.c: Ejemplo de una figura “Resumen del riesgo de sesgo”

8.7 EVALUACIONES RESUMEN DEL RIESGO DE SESGO

Tabla 8.7.a: Posible enfoque para las *evaluaciones resumen* del riesgo de sesgo para cada desenlace importante (entre los dominios) dentro y entre los estudios

8.8 INCORPORACIÓN DE LAS EVALUACIONES EN LOS ANÁLISIS

8.8.1 Introducción
8.8.2 Exploración de la repercusión del riesgo de sesgo
    8.8.2.1 Gráficos de los resultados según el riesgo de sesgo
    8.8.2.2 Estudios evaluados como riesgo poco claro de sesgo
    8.8.2.3 Metaregresión y comparaciones de subgrupos
8.8.3 Inclusión de las evaluaciones del “riesgo de sesgo” en los análisis

8.8.3.1 Posibles estrategias de análisis
8.8.4 Otros métodos para tratar con el riesgo de sesgo
8.8.4.1 Ponderación directa
8.8.4.2 Enfoques bayesianos

8.9 GENERACIÓN DE LA SECUENCIA

8.9.1 Justificación para las inquietudes acerca del sesgo
8.9.2 Evaluación del riesgo de sesgo con relación a la generación adecuada o inadecuada de la secuencia
8.9.2.1 Métodos adecuados de generación de la secuencia
8.9.2.2 Métodos inadecuados de generación de la secuencia
8.9.2.3 Métodos de generación de la secuencia con riesgo poco claro de sesgo

8.10 OCULTACIÓN DE LA SECUENCIA DE ASIGNACIÓN

8.10.1 Justificación para las inquietudes acerca del sesgo
8.10.2 Evaluación del riesgo de sesgo con relación a la ocultación adecuada o inadecuada de la asignación de la secuencia
8.10.2.1 Métodos adecuados de ocultación de la secuencia de asignación
Tabla 8.10.a: Criterios mínimos y ampliados para evaluar la ocultación de la secuencia de asignación como adecuada (bajo riesgo de sesgo)

8.11 CEGAMIENTO DE LOS PARTICIPANTES Y DE PERSONAL

8.11.1 Justificación para las inquietudes acerca del sesgo
Tabla 8.11.a: Una nota sobre el cegamiento en los ensayos clínicos
8.11.2 Evaluación del riesgo de sesgo con relación al cegamiento adecuado o inadecuado de los participantes y del personal

8.12 CEGAMIENTO DE LOS EVALUADORES

8.12.1 Justificación para las inquietudes acerca del sesgo
8.12.2 Valoración del riesgo de sesgo en relación al cegamiento adecuado o inadecuado de la evaluación de los resultados

8.13 DATOS DE DESENLACE INCOMPLETOS

8.13.1 Justificación para las inquietudes acerca del sesgo
8.13.2 Evaluación del riesgo de sesgo a partir de los datos de desenlace incompletos
8.13.2.1 Bajo riesgo de sesgo debido a datos de desenlace incompletos
8.13.2.2 Alto riesgo de sesgo debido a datos de desenlace incompletos
8.13.2.3 Intentos de solucionar los datos faltantes en las publicaciones: imputación

8.14 NOTIFICACIÓN SELECTIVA DE LOS DESENLACES

8.14.1 Justificación para las inquietudes acerca del sesgo
8.14.2 Evaluación del riesgo de sesgo de la notificación selectiva de los desenlaces

8.15 OTRAS AMENAZAS POTENCIALES A LA VALIDEZ

8.15.1 Justificación para las inquietudes acerca del sesgo
8.15.1.1 Riesgo de sesgo en diseños específicos
8.15.1.2 Desequilibrio inicial
8.15.1.3 Asignación en bloques en ensayos no cegados
8.15.1.4 Actividad diagnóstica diferencial
8.15.1.5 Otros ejemplos de sesgos potenciales
8.15.2 Evaluación del riesgo de sesgo de otras fuentes

8.16 INFORMACIÓN DEL CAPÍTULO

8.17 REFERENCIAS

CAPÍTULO 8: EVALUACIÓN DEL RIESGO DE SESGO EN LOS ESTUDIOS INCLUIDOS

Editors: Julian PT Higgins, Douglas G Altman y Jonathan AC Sterne del Grupo Cochrane de Métodos Estadísticos y el Grupo Cochrane de Métodos de Sesgo.

Puntos clave

- Los problemas con el diseño y la implementación de los estudios individuales de intervenciones en la atención sanitaria hacen surgir interrogantes acerca de la validez de sus hallazgos; la evidencia empírica apoya estas inquietudes.
- Una evaluación de la validez de los estudios incluidos en una revisión Cochrane debería hacer énfasis en el riesgo de sesgo de sus resultados, es decir, el riesgo de que los mismos sobrestimen o subestimen el verdadero efecto de la intervención.
- Hay numerosas herramientas disponibles para evaluar la calidad metodológica de los ensayos clínicos. No se recomienda el uso de escalas que producen una puntuación resumen.
- La Colaboración Cochrane recomienda una herramienta específica para evaluar el resigo de sesgo en cada estudio incluido. La misma incluye una descripción y una valoración para cada ítem en una tabla de “Riesgo de sesgo”, en la que cada ítem aborda un aspecto específico del estudio. La valoración para cada ítem incluye la respuesta a una pregunta, en la que las respuestas “Sí” indican un bajo riesgo de sesgo, “No” indican un alto riesgo de sesgo, e “Poco claro” indican falta de información o incertidumbre acerca del posible sesgo.
- Es posible crear gráficos de las evaluaciones del “Riesgo de sesgo” en RevMan.
- Para los ensayos de grupos paralelos las características de interés en una tabla estándar de “Riesgo de sesgo” de una revisión Cochrane son la generación de la secuencia, la ocultación de la secuencia de asignación, el cegamiento, los datos de resultado incompletos, la notificación selectiva de los resultados y otras posibles fuentes de sesgo.
- En este capítulo se proporcionan consideraciones detalladas para la evaluación de estas características.

8.1 Introducción

El grado en el cual una revisión Cochrane puede establecer conclusiones acerca de los efectos de una intervención depende de si los datos y resultados de los estudios incluidos son válidos. En particular, un metanálisis de estudios no válidos puede producir resultados erróneos con un intervalo de confianza estrecho alrededor de la estimación errónea del efecto de la intervención. Por lo tanto, la evaluación de la validez de los estudios incluidos es un componente fundamental de una revisión Cochrane y debería influir en el análisis, la interpretación y las conclusiones de la revisión.

Se puede considerar que la validez de un estudio tiene dos dimensiones. La primera dimensión es si el estudio formula la pregunta de investigación apropiada, lo que a menudo se describe como "validez externa", y su evaluación depende del objetivo para el cual se va a utilizar el estudio. La validez externa está íntimamente conectada con la generalizabilidad o la aplicabilidad de los hallazgos de un estudio, y se aborda en el Capítulo 12.

La segunda dimensión de la validez de un estudio se relaciona con el hecho de si el mismo responde a su pregunta de investigación "correctamente", es decir, de una manera libre de sesgo, lo que a menudo se describe como "validez interna" y que es el aspecto de la validez que se abordará en este capítulo. Como la mayoría de las revisiones Cochrane se centran en ensayos aleatorizados, esta parte se centrará en cómo evaluar la validez de este tipo de estudios. El Capítulo 13 aborda otros aspectos de la evaluación de los estudios no aleatorizados, y el Capítulo 14 incluye consideraciones adicionales para los efectos adversos. Las evaluaciones de la validez interna se citan frecuentemente como "evaluaciones de la calidad metodológica" o "evaluación de la calidad". Sin embargo, se evitará el término calidad por motivos que se explican más adelante. En la siguiente sección se define "sesgo" y se distingue de los conceptos relacionados de error al azar y calidad.

8.2 ¿Qué es sesgo?

8.2.1 "Sesgo" y "riesgo de sesgo"

Un sesgo es un error sistemático, o una desviación de la verdad, en los resultados o inferencias. Los sesgos pueden intervenir en cualquier dirección: sesgos diferentes pueden dar lugar a la subestimación o la sobrestimación del verdadero efecto de la intervención. Los sesgos pueden variar en cuanto a magnitud: algunos son pequeños (y triviales comparados con el efecto observado) y algunos son significativos (de manera que un hallazgo evidente se puede deber completamente a un sesgo). Incluso una fuente específica de sesgo puede variar en su dirección: el sesgo debido a un determinado defecto del diseño (p.ej. falta de ocultación de la asignación) puede dar lugar a la subestimación de un efecto en un estudio, y por otra parte a la sobrestimación en otro estudio. Habitualmente no es posible conocer hasta qué grado los sesgos han afectado los resultados de un estudio concreto, aunque existe evidencia empírica de buena calidad de que defectos específicos en el diseño, la realización y el análisis de los ensayos clínicos aleatorizados dan lugar a sesgo (ver Sección). Es más apropiado considerar el riesgo de sesgo debido a que los resultados de un estudio pueden de hecho no estar sesgados a pesar de los defectos metodológicos.

Las diferencias en los riesgos de sesgo pueden ayudar a explicar la variación en los resultados de los estudios incluidos en una revisión sistemática (es decir, pueden explicar la heterogeneidad de los resultados). Es más probable que los estudios más rigurosos produzcan resultados que estén más cerca de la verdad. El metanálisis de los resultados de estudios de validez variable pueden dar lugar a conclusiones positivas falsas (que concluyen erróneamente que una intervención es efectiva) si los estudios menos rigurosos están sesgados hacia la sobrestimación del efecto de una intervención. Dichos estudios también pueden dar lugar a conclusiones negativas falsas (que concluyen erróneamente que una

intervención no tiene efecto) si los estudios menos rigurosos están sesgados hacia la subestimación del efecto de una intervención (Detsky 1992).

Es importante evaluar el riesgo de sesgo en todos los estudios de una revisión independientemente de la variabilidad anticipada en los resultados o la validez de los estudios incluidos. Por ejemplo, los resultados pueden ser consistentes entre los estudios pero todos los estudios pueden presentar defectos. En este caso, las conclusiones de la revisión no deberían ser tan sólidas como las de una serie de estudios rigurosos que producen resultados consistentes acerca del efecto de una intervención. En una revisión Cochrane el proceso de evaluación se describe como la *evaluación del riesgo de sesgo en los estudios incluidos*. En la Sección 8.5 se describe una herramienta que se ha desarrollado e implementado en RevMan para este fin. El resto de este capítulo proporciona la justificación para esta herramienta, y explica cómo se deberían resumir e incorporar en el análisis las evaluaciones del sesgo (Secciones 8.6 a 8.8). Las secciones 8.9 a 8.14 proporcionan consideraciones básicas para ayudar a los revisores a utilizar la herramienta.

El sesgo no se debería confundir con la imprecisión. El sesgo se refiere al *error sistemático*, lo que significa que múltiples replicaciones del mismo estudio lograrían como promedio una respuesta errónea. La imprecisión se refiere al *error aleatorio*, lo que significa que múltiples replicaciones del estudio producirán diferentes estimaciones del efecto debido a la variación de la muestra, incluso si como promedio produjeran la respuesta correcta. Los resultados de los estudios más pequeños están sujetos a una mayor variación de la muestra, por lo que son menos precisos. La imprecisión se refleja en los intervalos de confianza alrededor de la estimación del efecto de la intervención de cada estudio y en la ponderación dada a los resultados de cada estudio en un metanálisis. A los resultados más precisos se les da una mayor ponderación.

8.2.2 “Riesgo de sesgo” y “calidad”

El sesgo se puede diferenciar de la calidad. La frase “evaluación de la calidad metodológica” se ha utilizado ampliamente en el contexto de los métodos de las revisiones sistemáticas para referirse a la evaluación critica de los estudios incluidos. El término indica una investigación del grado en el cual los autores del estudio realizaron su investigación con los estándares más altos posibles. Este *Manual* hace una distinción entre la evaluación de la calidad metodológica y la evaluación del riesgo de sesgo, y recomienda centrarse en la última. Los motivos para esta distinción incluyen:

1. La consideración clave en una revisión Cochrane es el grado en el cual se debería *creer* en los resultados de los estudios incluidos. La evaluación del riesgo de sesgo se dirige directamente a esta interrogante.
2. Es posible realizar un estudio con los estándares más altos posibles y que aún tenga un riesgo de sesgo importante. Por ejemplo, en muchas situaciones no es práctico ni posible cegar a los participantes o al personal del estudio al grupo de intervención. Es inapropiado y subjetivo describir todos estos estudios como de “calidad deficiente”, pero no significa que estén libres del sesgo derivado del conocimiento del estado de la intervención.
3. Es poco probable que algunos marcadores de calidad en la investigación médica como la obtención de la aprobación ética, realizar un cálculo del tamaño de muestra y presentar un estudio según la Declaración CONSORT (Moher 2001c), tengan implicaciones directas para el riesgo de sesgo.
4. El énfasis en el riesgo de sesgo resuelve la ambigüedad entre la calidad de la publicación y la calidad de la investigación subyacente (aunque no sortea el problema de tener que confiar en las publicaciones para evaluar la investigación subyacente).

A pesar de estas inquietudes acerca del término “calidad”, el término “calidad de la evidencia” se utiliza en las tablas “Resumen de los hallazgos” en las revisiones Cochrane para describir el grado en el cual es posible confiar en que una estimación del efecto está cerca del verdadero valor de un resultado entre los estudios, como se describe en el Capítulo 11 (Sección 11.5) y

en el Capítulo 12 (Sección 12.2). El riesgo de sesgo en los resultados de cada estudio que contribuye a una estimación del efecto es uno de los múltiples factores que se deberían considerar cuando se valora la calidad del cuerpo de la evidencia, como se define en este contexto.

8.2.3 Establecer la evidencia empírica de la presencia de sesgos

Los sesgos asociados con las características específicas de los estudios se pueden examinar mediante una técnica conocida a menudo como metaepidemiología (Naylor 1997, Sterne 2002). Un estudio metaepidemiológico analiza una colección de metanálisis, en cada uno de los cuales los estudios que lo constituyen se han clasificado según algunas características a nivel de estudio. Un antiguo ejemplo fue el estudio de ensayos clínicos con desenlaces dicotómicos incluidos en los metanálisis de la *Base de Datos del Grupo Cochrane de Embarazo y Parto* (*Cochrane Pregnancy and Childbirth Database)* (Schulz 1995b). Este estudio demostró que los ensayos en los cuales la aleatorización no se ocultó o no se describió adecuadamente produjeron estimaciones exageradas del efecto de la intervención, comparados con los ensayos que describieron una ocultación adecuada, y encontró una asociación similar (pero más pequeña) para los ensayos que no se describieron como doble ciego.

Un análisis sencillo de un estudio metaepidemiológico es calcular el "cociente de *odds ratios*" dentro de cada metanálisis (por ejemplo, los *odds ratios* de la intervención en los ensayos con ocultación inadecuada/incierta de la asignación divididos por los *odds ratio* de los ensayos con ocultación adecuada de la asignación). Estos cocientes de *odds ratios* luego se combinan entre los metanálisis en un metanálisis. Por lo tanto, dichos análisis también se conocen como "meta-metanálisis". En las secciones posteriores de este capítulo se cita evidencia empírica de sesgo a partir de los estudios metaepidemiológicos, cuando están disponibles, como parte del fundamento para evaluar cada dominio de posible sesgo.

8.3 Herramientas para evaluar la calidad y el riesgo de sesgo

8.3.1 Tipos de herramientas

Se han propuesto muchas herramientas para evaluar la calidad de los estudios con el fin de utilizarlos en el contexto de las revisiones sistemáticas u otros contextos. La mayoría de las herramientas son escalas en las que se les otorga una puntuación a varios componentes de calidad, la cual se combina para obtener una puntuación resumen; o listas de verificación, en la cual se formulan interrogantes específicos (Jüni 2001).

En 1995 Moher y cols. identificaron 25 escalas y nueve listas de verificación que se habían utilizado para evaluar la validez o la "calidad" de los ensayos aleatorizados (Moher 1995, Moher 1996). Estas escalas y listas de verificación incluyeron entre tres y 57 ítems, y se encontró que se tardaba entre diez y 45 minutos en completarlas para cada estudio. Casi todos los ítems en la herramienta se basaban en criterios indicados o "generalmente aceptados" que se mencionan en los libros de texto de ensayos clínicos. Muchas herramientas también contenían ítems que no se relacionaban directamente con la validez interna, como si se había realizado un cálculo del poder estadístico (un ítem que se relaciona más con la precisión de los resultados) o si los criterios de inclusión y exclusión se habían descrito claramente (un ítem que se relaciona más con la aplicabilidad que con la validez). Comparadas con las listas de verificación, es más probable que las escalas incluyan criterios que no se relacionan directamente con la validez interna.

La herramienta recomendada por la Colaboración para evaluar el riesgo de sesgo no es una escala ni una lista de verificación. Es una evaluación basada en dominios, en la cual las evaluaciones críticas se realizan de forma separada para diferentes dominios, descritos en la sección 8.5. Fue desarrollada entre 2005 y 2007 por un grupo de trabajo de metodólogos, editores y revisores. Debido a que es imposible conocer el grado de sesgo (o incluso el

verdadero riesgo de sesgo) en un estudio determinado, la posibilidad de validar cualquier herramienta propuesta es limitada. La evaluación más realista de la validez de un estudio puede incluir la subjetividad, por ejemplo, la evaluación de si la falta de cegamiento de los pacientes puede de forma plausible haber afectado la recurrencia de una afección grave como el cáncer.

8.3.2 Publicación frente a realización

Una dificultad clave en la evaluación del riesgo de sesgo o la calidad es el obstáculo debido a la publicación incompleta. Aunque se debería hacer énfasis en el riesgo de sesgo en el diseño real y la realización de un estudio, pudiera ser tentador recurrir a la evaluación de lo adecuado de la publicación. Muchas de las herramientas revisadas por Moher y cols. fueron proclives a confundir estos aspectos diferentes (Moher 1995). Además, las puntuaciones en las escalas se basaron a menudo en si algo se describió (como el señalar a cuántos participantes se asignaron) en lugar de si se hizo de forma apropiada en el estudio.

8.3.3 Escalas de calidad y revisiones Cochrane

No se recomienda explícitamente el uso de escalas para evaluar la calidad o el riesgo de sesgo en las revisiones Cochrane. Aunque el enfoque ofrece una sencillez atractiva, no está apoyado por evidencia empírica (Emerson 1990, Schulz 1995b). El cálculo de una puntuación resumen incluye inevitablemente la asignación de “ponderaciones” a diferentes ítems de una escala y es difícil justificar las ponderaciones asignadas. Además, las escalas han mostrado ser evaluaciones poco fiables de la validez (Jüni 1999) y es menos probable que sean transparentes para los usuarios de la revisión. Es preferible utilizar enfoques sencillos para evaluar la calidad que puedan describirse de forma completa (es decir, cómo se clasificó cada ensayo según cada criterio).

Una escala utilizada con frecuencia es la desarrollada por Jadad y cols. para los ensayos aleatorizados en la investigación del dolor (Jadad 1996). Se recomienda explícitamente no utilizar esta escala. Al tiempo que presenta los mismos problemas genéricos de las escalas, hace un fuerte énfasis en la publicación más que en la realización y no cubre uno los sesgos potenciales más importantes en los ensayos aleatorizados, a saber, la ocultación de la asignación (ver Sección 8.10.1).

8.3.4 Obtención de información para las evaluaciones del riesgo de sesgo

A pesar de las limitaciones de las publicaciones, la información acerca del diseño y la realización de los estudios se obtendrá de los informes publicados, incluidos los artículos de revistas, capítulos de libros, disertaciones, resúmenes de congresos y sitios web (incluidos registros de ensayos). Los protocolos publicados son una fuente especialmente valiosa de información cuando están disponibles. La extracción de la información de dichos informes se analiza en el Capítulo 7. Los formularios de obtención de datos deberían incluir espacio para extraer con detalle suficiente que permitar la implementación de la herramienta de la Colaboración “Riesgo de sesgo” (Sección 8.5). Cuando se extrae esta información, es especialmente conveniente registrar la fuente de cada pieza de información (incluida la localización exacta dentro del documento). Es útil probar los formularios de obtención de datos y las evaluaciones del riesgo de sesgo dentro del equipo de revisión con una muestra piloto de los artículos para asegurar que los criterios se apliquen de forma consistente, y que se alcance el consenso. Entre tres y seis artículos que, de ser posible, cubran un rango de bajo a alto riesgo de sesgo podrían proporcionar una muestra adecuada para esta prueba.

Los autores también deberían decidir si quienes evalúan el riesgo de sesgo estarán cegados a los nombres de los autores, las instituciones, la revista y los resultados de un estudio cuando se evalúan los métodos. Un estudio indicó que la evaluación cegada de las publicaciones podría producir calificaciones más bajas y más consistentes que las evaluaciones abiertas (Jadad 1996), mientras que otros indicaron un pequeño beneficio de las evaluaciones ciegas

(Berlin 1997, Kjaergard 2001). Las evaluaciones ciegas consumen mucho tiempo y es posible que no se puedan realizar cuando los estudios son bien conocidos por los revisores, y no todos los dominios del sesgo se pueden evaluar independientemente de los datos del desenlace. Además, el conocimiento de quién realizó un estudio permite en algunas ocasiones hacer suposiciones razonables acerca de cómo se realizó el estudio (aunque el revisor debería describir tales suposiciones). Los revisores deberían sopesar los beneficios potenciales contra los costes que conlleva cuando se decide cegar o no la evaluación de cierta información en las publicaciones de los estudios.

Los revisores con diferentes niveles de formación metodológica y experiencia pueden identificar diferentes fuentes de evidencia y hacer valoraciones diferentes acerca del riesgo de sesgo. Aunque los expertos en las áreas de contenido pueden tener opiniones preformadas que pueden influir en sus evaluaciones (Oxman 1993), pueden, no obstante, proporcionar evaluaciones más consistentes de la validez de los estudios que las personas sin experiencia en el contenido (Jadad 1996). Los expertos en contenido pueden tener ideas valiosas sobre la magnitud de los sesgos y los metodólogos experimentados pueden tener ideas valiosas de sesgos potenciales que en un inicio no son evidentes. Es conveniente que los revisores incluyan a expertos en contenido y metodólogos y aseguren que todos tengan una comprensión adecuada de los aspectos metodológicos relevantes.

A menudo los intentos de evaluar el sesgo se ven obstaculizados por la notificación incompleta de lo que sucedió durante la realización del estudio. Una opción para obtener la información faltante es establecer contacto con los investigadores de los estudios. Desafortunadamente los autores de contacto de las publicaciones de los ensayos pueden brindar respuestas exageradamente positivas. En una encuesta a 104 autores de ensayos, en la que se utilizaron preguntas directas acerca del cegamiento del personal del ensayo con categorías especificadas, el 43% respondió que los analistas de los datos en sus ensayos doble ciego estaban cegados, y el 19% respondió que quienes redactaron el texto estaban cegados (Haahr 2006). Esto es poco probable que sea cierto, ya que tales procedimientos se describieron en solamente el 3% y el 0% de los artículos publicados correspondientes, y se describen con muy poca frecuencia en otras publicaciones de ensayos.

Para reducir el riesgo de respuestas exageradamente positivas, los revisores deberían utilizar preguntas con respuesta abierta cuando les solicitan a los autores información acerca del diseño y la realización del estudio. Por ejemplo, para obtener información acerca del cegamiento pudiera ser apropiado una solicitud de la siguiente manera: “Por favor, describa todas las medidas utilizadas, si hubo alguna, para asegurar el cegamiento de los participantes del ensayo y del personal clave del ensayo, con respecto al conocimiento de qué intervención había recibido un participante”. Para obtener información acerca del proceso de aleatorización, podría ser apropiada una solicitud así: “¿Cómo decidió usted qué tratamiento recibiría el siguiente participante?” Luego se pueden formular preguntas más específicas para aclarar las dudas restantes.

8.4 Introducción a las fuentes de sesgo en los ensayos clínicos

La fiabilidad de los resultados de un ensayo aleatorizado depende del grado en el cual se hayan evitado las fuentes potenciales de sesgo. Una parte clave de una revisión es considerar el riesgo de sesgo en los resultados de cada estudio elegible. Una clasificación útil de los sesgos es la siguiente: sesgo de selección, sesgo de realización, sesgo de desgaste, sesgo de detección y sesgo de notificación. En esta sección se describen cada uno de estos sesgos y se presentan siete dominios correspondientes que se evalúan en la herramienta ‘Riesgo de sesgo’ de la Colaboración. Estos se resumen en la Tabla 8.4.a. Se describe la herramienta para valorar los siete dominios en la Sección 8.5. y se proporcionan consideraciones más detalladas de cada elemento en las Secciones 8.9 a 8.15.

8.4.1 Sesgo de selección

El sesgo de selección hace referencia a las diferencias sistemáticas entre las características iniciales de los grupos que se comparan. La única fortaleza de la asignación al azar es que, si se logra de forma exitosa, evita el sesgo de selección en la asignación de las intervenciones a los participantes. Este éxito depende de cumplir numerosos procesos interrelacionados. Se debería especificar una regla para la asignación de las intervenciones a los participantes, basada en algún proceso al azar (aleatorio). A este proceso se le llama generación de la secuencia. Además, se deberían tomar medidas para asegurar la implementación estricta de dicho esquema de asignación al azar para evitar el conocimiento previo de las asignaciones siguientes. A este proceso se le llama a menudo ocultación de la asignación, aunque se pudiera describir de forma más exacta como ocultación de la secuencia de asignación. Por lo tanto, un método adecuado para asignar las intervenciones sería utilizar una secuencia sencilla al azar (y por ello impredecible), y ocultar las asignaciones siguientes de quienes están involucrados en el reclutamiento en el ensayo.

Para todas las fuentes potenciales de sesgo, es importante considerar la magnitud y la dirección probables del sesgo. Por ejemplo, si todas las limitaciones metodológicas de los estudios pudieran sesgar los resultados hacia la falta de efecto y la evidencia indica que la intervención es efectiva, se podría concluir que la intervención es efectiva incluso con la presencia de estos sesgos potenciales.

8.4.2 Sesgo de realización

El sesgo de realización hace referencia a las diferencias sistemáticas entre grupos en la asistencia que se dispensa o bien en la exposición a otros factores además de las intervenciones de interés. Después del reclutamiento en el estudio, el cegamiento (o enmascaramiento) de los participantes y el personal del estudio puede reducir el riesgo de que el conocimiento de qué intervención se recibió, en lugar de la propia intervención, afecte los resultados. El cegamiento efectivo también puede asegurar que los grupos comparados reciban una cantidad similar de atención, tratamiento secundario e investigaciones diagnósticas. Sin embargo, el cegamiento no siempre es posible. Por ejemplo, habitualmente no es posible cegar a las personas con respecto a si se realizó o no una cirugía mayor.

8.4.3 Sesgo de detección

El sesgo de detección hace referencia a las diferencias sistemáticas entre grupos en la forma en qué los resultados fueron obtenidos. El cegamiento (o enmascaramiento) de los evaluadores puede reducir el riesgo de que conocer qué intervención se recibió, más que la propia intervención, afecta en la medida de los resultados. El cegamiento de los evaluadores puede ser especialmente importante para valorar variables de resultado subjetivas, como el nivel de dolor postoperatorio.

8.4.4 Sesgo de desgaste

El sesgo de desgaste hace referencia a las diferencias sistemáticas entre grupos en los abandonos del estudio. Estos comportan la notificación de los datos de desenlace incompletos. Hay dos razones para los abandonos o los datos de desenlace incompletos en los ensayos clínicos. Las exclusiones se refieren a situaciones en las que algunos participantes son omitidos de los análisis, a pesar de que los datos sobre sus resultados están disponibles. El desgaste hace referencia a situaciones en las que los datos sobre los resultados no están disponibles.

8.4.5 Sesgo de notificación

El sesgo de notificación hace referencia a las diferencias sistemáticas entre los resultados presentados y los no presentados. En un estudio publicado es más probable que se describan los análisis con diferencias estadísticamente significativas entre los grupos de intervención que los que presentan diferencias no significativas. Este tipo de "sesgo de publicación dentro del estudio" se conoce habitualmente como sesgo de notificación selectivo y puede ser uno de los sesgos más importantes que afecten los resultados de los estudios individuales (Chan 2005).

8.4.6 Otros sesgos

Además, hay otras fuentes de sesgo que son relevantes sólo en ciertas circunstancias. Algunas se pueden encontrar solamente en diseños determinados de ensayos (p.ej. el efecto de arrastre [*carry-over*] en los ensayos cruzados [*crossover*] y el sesgo de reclutamiento en los ensayos aleatorizados grupales); algunas se pueden encontrar en un amplio espectro de ensayos, pero sólo en circunstancias específicas (p.ej. contaminación, en la cual las intervenciones del grupo experimental y el control "se mezclan", por ejemplo si los participantes combinan sus fármacos) y puede haber fuentes de sesgo que solamente se encuentran en un ámbito clínico determinado.

Tabla 8.4.a: Un esquema de clasificación frecuente para el sesgo

Tipo de sesgo	Descripción	Dominios relevantes en la herramienta de la Colaboración "Riesgo de sesgo"
Sesgo de selección.	Diferencias sistemáticas entre las características iniciales de los grupos que se comparan.	Generación de la secuencia. Ocultación de la asignación.
Sesgo de realización.	Diferencias sistemáticas entre los grupos en cuanto a la atención que se proporciona, o en la exposición a factores diferentes de la intervención de interés.	Cegamiento de los participantes y del personal. Otras amenazas potenciales a la validez.
Sesgo de detección.	Diferencias sistemáticas entre los grupos en cómo se determinaron los resultados.	Cegamiento de los evaluadores de resultado. Otras amenazas potenciales a la validez.
Sesgo de desgaste.	Diferencias sistemáticas entre los grupos en los abandonos de un estudio.	Datos de resultado incompletos.
Sesgo de notificación.	Diferencias sistemáticas entre los hallazgos presentados y no presentados.	Notificación selectiva de los resultados (ver también Capítulo 10).

8.5 Herramienta de la Colaboración Cochrane para evaluar el riesgo de sesgo

8.5.1 Generalidades

Esta sección describe el enfoque recomendado para evaluar el riesgo de sesgo en los estudios incluidos en las revisiones Cochrane. Es una herramienta en dos partes que aborda los seis dominios específicos discutidos en las Secciones 8.9 a 8.14 (a saber, generación de la secuencia, ocultación de la asignación, cegamiento, datos de resultado incompletos, notificación selectiva de los resultados y "otros aspectos"). La herramienta se resume en la Tabla 8.5.a. Cabe destacar que la herramienta se revisó a finales de 2010 después de un proyecto de evaluación. Los cambios realizados en aquel momento se resumen en la Table 8.5.b.

Cada dominio incluye uno o más ítems específicos en una tabla de "Riesgo de sesgo". Dentro de cada ítem, la primera parte de la herramienta incluye la descripción de qué se describió que sucedió en el estudio. La segunda parte de la herramienta incluye la asignación de una valoración con relación al riesgo de sesgo para ese ítem. Esto se logra al asignar una valoración de 'Bajo riesgo' de sesgo, 'Alto riesgo' de sesgo o 'Riesgo poco claro' de sesgo.
Los dominios generación de la secuencia, ocultación de la asignación y notificación selectiva de los resultados se deberían abordar en la herramienta con un ítem único para cada estudio. Para el cegamiento y para los datos de resultado incompletos es posible utilizar dos o más ítems porque generalmente es necesario realizar evaluaciones por separado para cada resultado (o para el mismo resultado en diferentes puntos temporales). Los revisores deberían tratar de limitar el número de ítems utilizados agrupando los resultados, por ejemplo, como resultados "subjetivos" u "objetivos" a los fines de la evaluación del cegamiento; o como "notificado por el paciente a los seis meses" o "notificado por el paciente a los 12 meses" para los datos de resultado incompletos. Se aplicará el mismo agrupamiento de los resultados a cada estudio de la revisión. El dominio final ("otras fuentes de sesgo") se puede evaluar como un ítem único para los estudios como un todo (por defecto en RevMan). Sin embargo, se recomienda que se utilicen múltiples ítems preespecificados para abordar otros riesgos específicos de sesgo. Dichos ítems especificados por el autor pueden ser para estudios como un todo o para resultados individuales (o agrupados) dentro de cada estudio.

Tabla 8.5.a: Herramienta de la Colaboración Cochrane para evaluar el riego de sesgo

Dominio	Descripción	Valoración de los revisores
Sesgo de selección.		
Generación de la secuencia.	Describir el método utilizado para generar la secuencia de asignación con detalle suficiente para permitir una evaluación de si la misma produjo grupos comparables.	Sesgo de selección (asignación sesgada a las intervenciones) a causa de una generación inadecauda de la secuencia de aleatorización.
Ocultamiento de la asignación.	Describir el método utilizado para ocultar la secuencia de asignación con detalle suficiente para determinar si las asignaciones a la intervención se podían prever antes o durante el reclutamiento.	Sesgo de selección (asignación sesgada a las intervenciones) a causa de una ocultación inadecuada de las asignaciones antes de asignarlas.
Sesgo de realización.		
Cegamiento de los participantes y del personal *Se debería evaluar cada resultado principal (o cada clase de resultado).*	Describir todas las medidas utilizadas, si se utilizó alguna, para cegar a los participantes y al personal del estudio al conocimiento de qué intervención recibió un participante. Proporcionar cualquier información con respecto a si el cegamiento propuesto fue efectivo.	Sesgo de realización a causa del conocimiento por parte de los participantes y del personal durante el estudio de las intervenciones asignadas.
Sesgo de detección.		
Cegamiento de los evaluadores del resultado *Se debería evaluar cada resultado principal (o cada clase de resultado).*	Describir todas las medidas utilizadas, si se utilizó alguna, para cegar a los evaluadores del resultado del estudio al conocimiento de qué intervención recibió un participante. Proporcionar cualquier información con respecto a si el cegamiento propuesto fue efectivo.	Sesgo de detección a causa del conocimiento por parte de los evaluadores de los resultados de las intervenciones asignadas.
Sesgo de desgaste.		
Datos de resultado incompletos *Se debería evaluar cada resultado principal (o cada clase de resultado).*	Describir la compleción de los datos de resultado para cada resultado principal, incluidos los abandonos y las exclusiones del análisis. Señalar si se describieron las los abandonos y las exclusiones, los números en cada grupo de intervención (comparados con el total de participantes asignados al azar), los motivos de las deserciones/exclusiones cuando se detallaron, y cualquier reinclusión en los análisis realizada por los revisores.	Sesgo de desgaste a causa de la cantidad, la naturaleza o el manejo de los datos de resultado incompletos.
Sesgo de notificación.		

Notificación selectiva de los resultados. *Se debería evaluar cada resultado principal (o cada clase de resultado).*	Señalar cómo los revisores examinaron la posibilidad de la notificación selectiva de los resultados, y qué encontraron.	Sesgo de notificación a causa de la notificación selectiva de los resultados.
Otros sesgos.		
Otras fuentes de sesgo.	Señalar alguna inquietud importante acerca del sesgo no abordada en los otros dominios del instrumento. Si en el protocolo de la revisión se prespecificaron preguntas/ítems particulares, se deberían proporcionar las respuestas para cada pregunta/ítem.	Sesgo debido a otros problemas no abordados en los apartados anteriores.

Table 8.5.b: Diferencias entre la herramienta 'Riesgo de sesgo' descrita en el Manual versiones 5.0.1/5.0.2 y la herramienta 'Riesgo de sesgo' revisada descrita en el Manual versión 5.1.0. (esta versión)

Separación del cegamiento.	En la versión anterior, los sesgos relacionados con el cegamiento de los participantes, el personal y los evaluadores se evaluaban en un dominio único (aunque se podían valorar de forma separada para resultados diferentes). En la herramienta revisada, el sesgo relacionado con el cegamiento de los participantes y del personal se valora en un dominio de forma separada del sesgo relacionada con la valoración de los resultados.
Carácter de la valoración	Las valoraciones ahora se expresan de una forma más sencilla como 'Bajo riesgo', 'Alto riesgo' o 'Riesgo poco claro' de sesgo. Se han retirado las preguntas, así como las respuestas 'Sí' indicando bajo riesgo de sesgo y 'No' indicando alto riesgo de sesgo.
Modificaciones menores	Algunos ítems se han renombrado en RevMan con la eliminación de las valoraciones basadas en las preguntas: *¿Generación adecuada de la secuencia?* se convierte en *Generación de la secuencia aleatoria* *¿Ocultación de la asignación?* se convierte en *Ocultación de la asignación* *¿Cegamiento?* se convierte en *Cegamiento de los participantes y del personal y Cegamiento de los evaluadores de los resultados.* *¿Se abordaron los datos de resultado incompletos?* se convierte en *Datos de resultado incompletos.* *¿Libre de notificación selectiva?* se convierte en *Notificación selectiva.* *¿Libre de otros sesgos? se convierte en Otros sesgos.*
Inserción de las categorías del sesgo	La herramienta revisada clarifica la categoría del sesgo según el dominio al que pertenezca: sesgo de selección (generación de la secuencia aleatoria y ocultación de la asignación), sesgo de realización (cegamiento de los participantes y del personal), sesgo de detección (cegamiento de los evaluadores), sesgo de desgaste (datos de resultado incompletos) y sesgo de notificación (notificación selectiva) y otros sesgos.
Reconsideración de ítems adecuados para el apartado 'otros sesgos', incluyendo la interrupción anticipada de un ensayo.	Las orientaciones sobre el dominio 'otros sesgos' se ha editado para enfatizar el hecho de que los ítems adicionales solamente se deberían utilizar excepcionalmente y que estos ítems deberían referirse con aspectos que podrían llevar al sesgo directamente. En particular, la mención de la interrupción prematura de un ensayo se ha eliminado, porque (i) la evidencia simulada sugiere que la inclusión de los ensayos interrumpidos prematuramente no conllevará un sesgo sustancial y (ii) la exclusión de ensayos interrumpidos prematuramente tiene el potencial de sesgar el metanálisis hace el efecto nulo (así como comportar una pérdidad de precisión).

8.5.2 Ayuda para la evaluación

La ayuda para la evaluación proporciona un resumen breve sobre qué valoraciones sobre el riesgo de sesgo pueden realizarse y tiene el objetivo de asegurar la transparencia sobre la forma en qué se obtienen estas valoraciones. Para un estudio específico, la información para la evaluación a menudo se obtendrá de un único estudio publicado, pero se puede obtener de una mezcla de estudios, protocolos, comentarios publicados sobre un artículo y contactos con los investigadores. De ser apropiado, la descripción debería incluir citas literales de los documentos o la correspondencia. Alternativamente, o de forma adicional, es posible incluir un resumen de hechos conocidos, o un comentario de los revisores. En particular, se debería incluir otra información que haya influido en cualquier evaluación realizada (como el conocimiento de otros estudios realizados por los mismos investigadores). Una construcción útil para complementar una cita ambigua es señalar "Probablemente realizado" o "Probablemente no realizado", siempre que existan motivos para tales afirmaciones. Cuando no haya información disponible a partir de la cual realizar una evaluación, se debería señalar de forma explícita. En la Tabla 8.5.c se proporcionan ejemplos del formato propuesto para la descripción.

Tabla 8.5.c: Ejemplos de ayudas para la evaluación de la generación de la secuencia

Generación de la secuencia.	Comentario: No se proporciona información.
Generación de la secuencia.	Cita: "los pacientes se asignaron al azar".
Generación de la secuencia.	Cita: "los pacientes se asignaron al azar". Comentario: Probablemente realizada, ya que las primeras publicaciones de los mismos investigadores describieron claramente el uso de secuencias aleatorias (Cartwright 1980).
Generación de la secuencia.	Cita: "los pacientes se asignaron al azar". Comentario: Probablemente no realizada, ya que un ensayo similar de estos investigadores incluyó la misma frase, aunque utilizaron asignación alterna (Winrow 1983).
Generación de la secuencia.	Cita (de la publicacion): "los pacientes se asignaron al azar". Cita (de la correspondencia): "La asignación al azar se realizó según el día del tratamiento". Comentario: No aleatorizado.

8.5.3 La evaluación

Las evaluaciones de los revisores deberían clasificarse en 'Bajo riesgo' de sesgo, 'Alto riesgo' de sesgo y 'Riesgo poco claro' de sesgo. Las evaluaciones deberían considerar el riesgo del sesgo material en lugar de cualquier sesgo. Se define el 'riesgo material' como cualquier sesgo de magnitud suficiente para tener un impacto notable en los resultados o en las conclusiones del ensayo, reconociendo que la subjetividad está involucrada en cada valoración.

La Tabla 8.5.d proporciona los criterios para realizar las evaluaciones acerca del riesgo de sesgo de cada uno de los seis dominios de la herramienta. Si no se describen detalles suficientes acerca de qué sucedió en el estudio, habitualmente se evaluará como riesgo de sesgo "Poco claro". También se debería evaluar como "Poco claro" si se conoce lo que sucedió en el estudio pero se desconoce el riesgo de sesgo; o si el ítem disponible no es relevante para el estudio (especialmente para evaluar el cegamiento y los datos de resultado incompletos, en los que el resultado que se evalúa mediante el ítem no se midió en el estudio).

Table 8.5.d: Criterios para evaluar el riesgo de sesgo en la herramienta de evaluación "Riesgo de sesgo"

GENERACIÓN ALEATORIA DE LA SECUENCIA Sesgo de selección (asignación sesgada a las intervenciones) a causa de una generación inadecuada de la secuencia de aleatorización.	
Criterios para una evaluación de 'Bajo riesgo' de sesgo	Los investigadores describen un componente aleatorio en el proceso de generación de la secuencia como: ▪ Se refieren a una tabla de números aleatorios; ▪ Uso de un generador de números aleatorios por ordenador; ▪ Lanzamiento de una moneda; ▪ Barajar cartas o sobres; ▪ Lanzar los dados; ▪ Sorteo de tarjetas; ▪ Minimización*. *La minimización se puede implementar sin un elemento aleatorio, lo que se considera equivalente a que sea aleatorio.
Criterios para una evaluación de 'Alto riesgo' de sesgo	Los investigadores describen un componente no aleatorio en el proceso de generación de la secuencia. Habitualmente la descripción involucra algún enfoque sistemático y no aleatorio, por ejemplo: ▪ Secuencia generada mediante la fecha de nacimiento par o impar; ▪ Secuencia generada mediante alguna regla según la fecha (o el día) de ingreso; ▪ Secuencia generada mediante alguna regla según la historia clínica del hospital o el consultorio. Otros enfoques no aleatorios se utilizan con mucha menor frecuencia que los enfoques sistemáticos mencionados anteriormente y tienden a ser obvios. Habitualmente incluyen la evaluación o algún método de categorización no aleatoria de los participantes, por ejemplo: ▪ Asignación según el criterio del médico; ▪ Asignación según la preferencia del participante; ▪ Asignación según los resultados de una prueba de laboratorio o una serie de pruebas; ▪ Asignación según la disponibilidad de la intervención.
Criterios para una evaluación de 'Riesgo poco claro' de sesgo	No hay información suficiente acerca del proceso de generación de la secuencia para permitir una evaluación de "Bajo riesgo" o "Alto riesgo".
OCULTACIÓN DE LA ASIGNACIÓN Sesgo de selección (asignación sesgada a las intervenciones) a causa de una ocultación inadecuada de las asignaciones.	
Criterios para una evaluación de 'Bajo riesgo' de sesgo	Los participantes y los investigadores que reclutaron a los participantes no podían prever la asignación debido a que uno de los métodos siguientes u otro equivalente se utilizaron para ocultar la asignación: ▪ Asignación central (incluida la asignación al azar por teléfono, basada en la web y controlada por la farmacia); ▪ Envases del fármaco, numerados de forma secuencial con apariencia idéntica; ▪ Sobres cerrados, oscuros y numerados de forma secuencial.
Criterios para una evaluación de 'Alto riesgo' de sesgo	Los participantes o los investigadores que reclutaron a los participantes podían prever las asignaciones y por lo tanto introducir sesgo de selección, por ejemplo, asignación según: ▪ Uso de un esquema de asignación al azar abierto (p.ej. una lista de

	números aleatorios); ▪ Se utilizaron sobres de asignación sin una protección adecuada (p.ej. si los sobres no estaban cerrados, no eran oscuros ni estaban numerados de forma secuencial); ▪ Alternancia o rotación; ▪ Fecha de nacimiento; ▪ Número de historia clínica; ▪ Cualquier otro procedimiento explícitamente no ocultado.
Criterios para una evaluación de 'Riesgo poco claro' de sesgo	No hubo información suficiente para permitir una evaluación de "Bajo riesgo" o "Alto riesgo". Éste es habitualmente el caso si el método de ocultación no se describe o no se describe con detalle suficiente para permitir una evaluación definitiva, por ejemplo, si se describe el uso de sobres de asignación, pero es incierto si los sobres eran oscuros y estaban numerados de forma secuencial y cerrados.
CEGAMIENTO DE LOS PARTICIPANTES Y DEL PERSONAL Sesgo de realización a causa del conocimiento de las intervenciones asignadas por parte de los participantes y del personal durante el estudio.	
Criterios para una evaluación de 'Bajo riesgo' de sesgo	Cualquiera de los siguientes: ▪ Ningún cegamiento, pero los revisores consideran que no es probable que el resultado esté influido por la falta de cegamiento; ▪ Se aseguró el cegamiento de los participantes y el personal clave del estudio, y es poco probable que se haya roto el cegamiento.
Criterios para una evaluación de 'Alto riesgo' de sesgo	Cualquiera de los siguientes: ▪ Ningún cegamiento o cegamiento incompleto, y es probable que el resultado y la medición del resultado estén influidos por la falta de cegamiento; ▪ Se intentó el cegamiento de los participantes y el personal clave del estudio, pero es probable que se haya roto el cegamiento.
Criterios para una evaluación de 'Riesgo poco claro' de sesgo	Cualquiera de los siguientes: ▪ No hubo información suficiente para permitir una evaluación de "Bajo riesgo" o "Alto riesgo". ▪ El estudio no abordó este resultado.
CEGAMIENTO DE LOS EVALUADORES DE LOS RESULTADOS Sesgo de detección a causa del conocimiento de las intervenciones asignadas por parte de los evaluadores.	
Criterios para una evaluación de 'Bajo riesgo' de sesgo	Cualquiera de los siguientes: ▪ No hay cegamiento de la evaluación de resultados, pero los revisores creen que la medida del resultado no es probable que esté influenciada por la falta de cegamiento; ▪ Se ha asegurado el cegamiento de la evaluación de los resultados y es improbable que se haya roto el cegamiento.
Criterios para una evaluación de 'Alto riesgo' de sesgo	Cualquiera de los siguientes: ▪ No hay cegamiento de la evaluación de resultados y es probable que la medida del resultado esté influenciada por la falta de cegamiento; ▪ Se ha realizado la valoración del cegamiento, pero es probable que el cegamiento se haya roto y que la medida del resultado esté influenciada por la falta de cegamiento.
Criterios para una evaluación de 'Riesgo poco claro' de sesgo	Cualquiera de los siguientes: ▪ No hay suficiente información para permitir una evaluación de 'Bajo riesgo' o 'Alto riesgo'; ▪ El estudio no abordó este resultado.
DATOS DE RESULTADO INCOMPLETOS Sesgo de desgaste debido a la cantidad de datos de resultado incompletos, su	

naturaleza o su manejo.	
Criterios para una evaluación de 'Bajo riesgo' de sesgo	Cualquiera de los siguientes: ▪ No hay datos de resultado faltantes; ▪ Es poco probable que los motivos causantes de los datos de resultado faltantes se relacionen con el resultado verdadero (para los datos de supervivencia, es poco probable que la censura de los datos haya introducido sesgo); ▪ Datos de resultado faltantes equilibrados con respecto a los números entre los grupos de intervención, con motivos similares para los datos de resultado faltantes entre los grupos; ▪ Para los datos de resultado dicotómicos, la proporción de resultados faltantes comparados con el riesgo de evento observado no es suficiente para tener una repercusión clínicamente importante sobre la estimación del efecto de la intervención; ▪ Para los datos de resultado continuos, el posible tamaño del efecto (diferencia de medias o diferencia de medias estandarizada) entre los resultados faltantes no es suficiente para tener una repercusión clínicamente relevante sobre el tamaño del efecto observado; ▪ Los datos faltantes se imputaron mediante métodos apropiados.
Criterios para una evaluación de 'Alto riesgo' de sesgo	Cualquiera de los siguientes: ▪ Es probable que las razones que causan los datos faltantes estén relacionados con el resultado verdadero, con desequilibrio en los números o en los motivos para los datos faltantes entre los grupos de intervención; ▪ Para los datos de resultado dicotómicos, la proporción de resultados faltantes comparados con el riesgo de evento observado es suficiente para inducir un sesgo clínicamente relevante en la estimación del efecto de la intervención; ▪ Para los datos de resultado continuos, el posible tamaño del efecto (diferencia de medias o diferencia de medias estandarizada) entre los resultados faltantes es suficiente para inducir un sesgo clínicamente relevante en el tamaño del efecto observado; ▪ El análisis "como se trató" se realizó de forma significativamente diferente de la intervención recibida con respecto a la asignada mediante la asignación aleatoria; ▪ Aplicación posiblemente inapropiada de la imputación simple.
Criterios para una evaluación de 'Riesgo poco claro' de sesgo	Cualquiera de los siguientes: ▪ No hay datos suficientes sobre las pérdidas/exclusiones para permitir una evaluación de 'Bajo riesgo' o 'Alto riesgo' (por ejemplo, no se señala el número de participantes asignados al azar, no se proporcionan los motivos de los datos faltantes); ▪ El estudio no abordó este resultado.
NOTIFICACIÓN SELECTIVA DE LOS RESULTADOS Sesgo de notificación a causa de una notificación selectiva incompleta.	
Criterios para una evaluación de 'Bajo riesgo' de sesgo	Cualquiera de los siguientes: ▪ El protocolo del estudio está disponible y todos los resultados preespecificados (primarios y secundarios) del estudio que son de interés para la revisión se describieron de una manera preespecificada; ▪ El protocolo del estudio no está disponible pero está claro que las publicaciones incluyen todos los resultados esperados, incluidos los que se preespecificaron (puede ser poco frecuente la presencia de texto convincente de esta naturaleza).
Criterios para una evaluación de 'Alto riesgo' de sesgo	Cualquiera de los siguientes: ▪ No se describieron todos los resultados primarios del estudio preespecificados;

	▪ Uno o más resultados primarios se describieron con mediciones, métodos de análisis o subgrupos de datos (p.ej. subescalas) que no se prespecificaron; ▪ Uno o más resultados primarios no se prespecificaron (a menos que se proporcione una clara justificación para detallarlos, como un efecto adverso inesperado); ▪ Uno o más resultados preespecificados de interés para la revisión se describieron de forma incompleta, por lo que no fue posible introducirlos en un metanálisis; ▪ La publicación del estudio no incluyó resultados para un resultado clave que era de esperar que se describiera para esta clase de estudios.
Criterios para una evaluación de 'Riesgo poco claro' de sesgo	No hubo información suficiente para permitir una evaluación de "Bajo riesgo" o "Alto riesgo". Es probable que la mayoría de los estudios se incluya en esta categoría.
OTROS SESGOS Sesgos debidos a otros problemas no abordados en los apartados anteriores.	
Criterios para una evaluación de 'Bajo riesgo' de sesgo	El estudio parece estar libre de otras fuentes de sesgo.
Criterios para una evaluación de 'Alto riesgo' de sesgo	Hay al menos un riesgo importante de sesgo. Por ejemplo, el estudio: ▪ Tiene una fuente potencial de sesgo relacionada con el diseño específico utilizado en el estudio; o ▪ Tuvo un desequilibrio inicial extremo; o ▪ Se ha señalado la presencia de fraude; o ▪ Tiene algún otro problema.
Criterios para una evaluación de 'Riesgo poco claro' de sesgo	Puede haber riesgo de sesgo, pero no hay: ▪ Información suficiente para evaluar si existe un riesgo importante de sesgo; o ▪ Justificación o pruebas suficientes de que un problema identificado introducirá sesgo.

8.6. Presentación de las evaluaciones del riesgo de sesgo

En RevMan está disponible una tabla de "Riesgo de sesgo" para su inclusión en una revisión Cochrane como parte de la tabla "Características de los estudios incluidos". Para cada ítem, la evaluación ("Bajo riesgo" de sesgo, "Alto riesgo" de sesgo o "Poco claro" riesgo de sesgo) es seguida por un cuadro de texto que proporciona una descripción del diseño, la realización o las observaciones que fundamentan la evaluación. La figura 8.6.a proporciona un ejemplo de cómo se pueden ver estos gráficos. Si el cuadro de texto se deja vacío, y la evaluación se deja como "Poco claro", entonces se omitirá el ítem en la tabla "Riesgo de sesgo" para el estudio cuando se publique en la *CDSR*.

Las consideraciones para la presentación de las evaluaciones del "Riesgo de sesgo" en el text de revisión se discuten en el Capítulo 4 (Sección 4.5) (bajo el subtítulo de los resultados "Riesgo de sesgo en los estudios incluidos" y el subtítulo de la discusión "Calidad de la evidencia".

Se pueden generar dos figuras para su inclusión en una revisión publicada utilizando RevMan. En primer lugar, el "Gráfico del riesgo de sesgo" ilustra la proporción de estudios con cada una de sus evaluaciones ("Bajo riesgo", "Alto riesgo", "Riesgo poco claro"). En segundo lugar, la figura "Resumen del riesgo de sesgo" presenta todas las evaluaciones en una tabulación cruzada por entrada (ver Figura 8.6.c).

Una versión alternativa de la primera figura (el "Gráfico del riesgo de sesgo") sería restringir la información a los estudios de un determinado metanálisis importante, y representar la proporción de la información (en lugar de la proporción de los estudios) de bajo, alto y poco claro riesgo de sesgo. La proporción de la información se podría medir por la suma de los pesos adjudicados a los estudios en el metanálisis. De todas formas, estos gráficos no pueden ser realizados en RevMan por ahora.

Figura 8.6.a: Ejemplo de una tabla de "Riesgo de sesgo" para un estudio único (ficticio)

Item	Evaluación	Apoyo para la valoración
Generación de la secuencia aleatorizada (sesgo de selección)	Bajo riesgo	Cita: "los pacientes se asignaron de forma aleatoria" Comentario: probablemente realizada, ya que las primeras publicaciones de los mismos investigadores describieron claramente el uso de secuencias aleatorias (Cartwright 1980).
Ocultación de la asignación (sesgo de selección)	Alto riesgo	Cita: "… mediante una tabla de números aleatorios." Comentario: probablemente no realizada.
Cegamiento de los participantes y del personal (sesgo de realización)	Bajo riesgo	Cita: "doble ciego, doble simulación", "Los comprimidos con dosis altas y bajas no eran distinguibles en todos sus aspectos en cuanto a su apariencia externa. Para cada fármaco hubo un placebo idéntico pareado (el éxito del cegamiento se evaluó al examinar los fármacos antes de la distribución)." Comentario: probablemente realizado.
Cegamiento de los evaluadores (sesgo de detección) (resultados notificados por el paciente)	Bajo riesgo	Cita: "doble ciego" Comentario: probablemente realizado.

Cegamiento de los evaluadores (sesgo de detección) (mortalidad)	Bajo riesgo	Obtenido de la historia clínica. Los revisores no creen que esto introduzca sesgo.
Manejo de los datos de resultado incompletos (sesgo de desgaste) (resultados a corto plazo) [2 a 6 semanas]	Alto riesgo	Cuatro semanas: 17/110 faltantes del grupo de intervención (nueve debido a "falta de eficacia"); 7/113 faltantes del grupo control (dos debido a "falta de eficacia").
Manejo de los datos de resultado incompletos (sesgo de desgaste) (resultados a más largo plazo) [> 6 semanas]	Alto riesgo	12 semanas: 31/110 faltantes del grupo intervención; 18/113 faltantes de grupo control. Los motivos difieren entre los grupos.
Notificación selectiva (sesgo de notificación)	Alto riesgo	Tres escalas de clasificación para la cognición descritas en "Métodos", pero sólo se describió una (con resultados estadísticamente significativos).

Figura 8.6.b: Ejemplo de una figura "Gráfico del riesgo de sesgo"

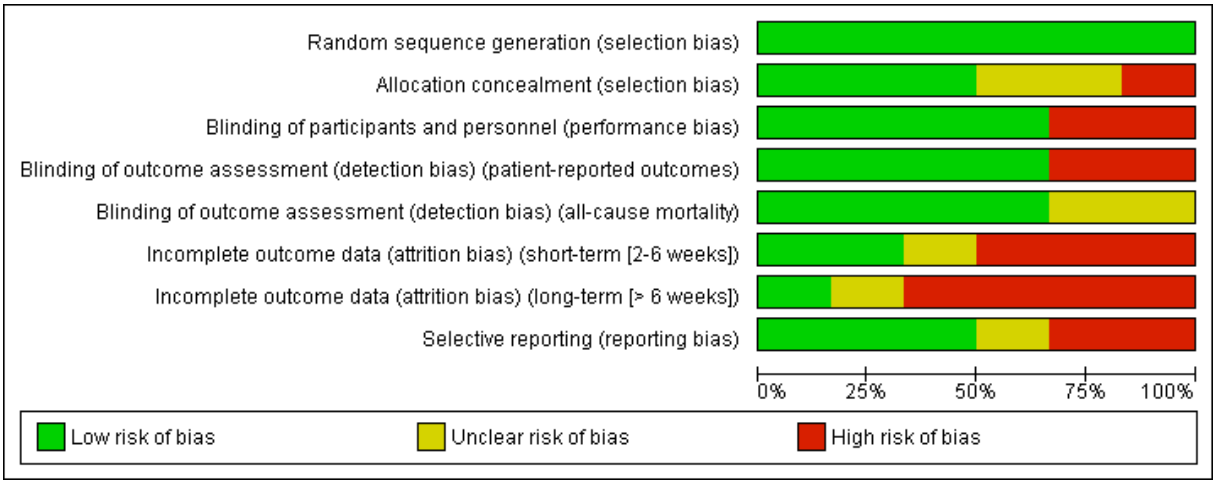
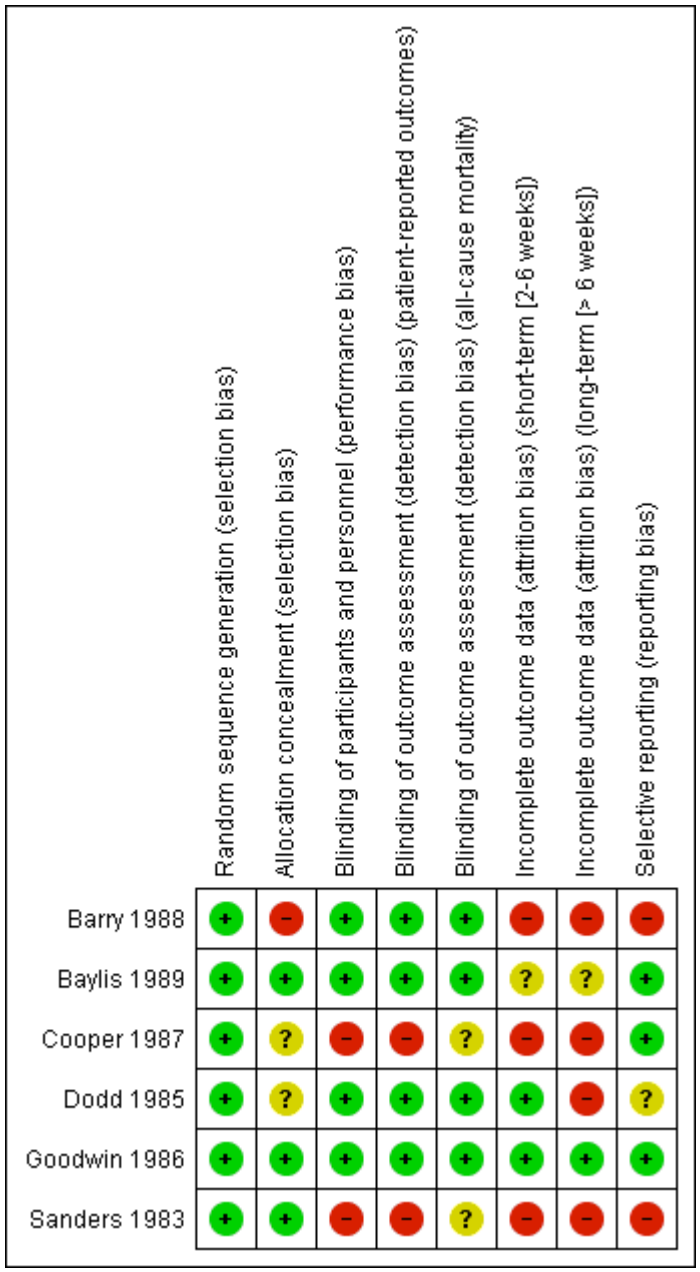


Figura 8.6.c: Ejemplo de una figura "Resumen del riesgo de sesgo"



8.7 Evaluaciones resumen del riesgo de sesgo

La herramienta recomendada por la Colaboración para evaluar el riesgo de sesgo en los estudios incluidos incluye la evaluación y presentación de dominios individuales, como la ocultación de la asignación y el cegamiento. Para establecer conclusiones acerca del riesgo general de sesgo para un desenlace es necesario resumir dicho riesgo. No se recomienda el uso de escalas (en las cuales se suman las puntuaciones de varios ítems para producir un total) por los motivos mencionados en la Sección 8.3.1.

Sin embargo, cualquier evaluación del riesgo general de sesgo incluye considerar la importancia relativa de los diferentes dominios. Un revisor tendrá que evaluar qué dominios son más importantes en la presente revisión. Por ejemplo, para desenlaces muy subjetivos como el dolor, los revisores pueden decidir que el cegamiento de los participantes es

fundamental. Se debería señalar explícitamente cómo se alcanzaron dichas evaluaciones, las cuales se deberían informar por:

- Evidencia empírica de sesgo: Las Secciones 8.5 a 8.15 resumen la evidencia empírica de la asociación entre los dominios como la ocultación de la asignación y el cegamiento y las estimaciones de la magnitud del efecto. Sin embargo, la base de la evidencia aún es incompleta.

- Dirección probable del sesgo: La evidencia empírica disponible indica que la imposibilidad de cumplir con la mayoría de los criterios, como la ocultación adecuada de la asignación, se asocia con sobrestimaciones del efecto. Si la dirección probable del sesgo para un dominio es tal que se realizará una subestimación del efecto (sesgado hacia la nulidad) entonces, siempre que la revisión demuestre un efecto importante de la intervención, dicho dominio podría tener una menor importancia.

- Magnitud probable del sesgo: La magnitud probable del sesgo asociado con cualquier dominio puede variar. Por ejemplo, es probable que la magnitud del sesgo asociado con el cegamiento inadecuado de los participantes sea mayor para los desenlaces más subjetivos. La base de la evidencia empírica puede proporcionar cierta indicación sobre la probable magnitud del sesgo (ver anteriormente), pero todavía no proporciona información clara sobre las situaciones particulares en las cuales los sesgos pueden ser grandes o pequeños. Sin embargo, es posible considerar la magnitud probable del sesgo con relación a la estimación de la magnitud del efecto. Por ejemplo, la ocultación inadecuada de la secuencia de asignación y una estimación pequeña del efecto podría reducir significativamente la confianza en la estimación, mientras que es posible que las deficiencias menores en cómo se abordaron los datos de resultado incompletos no reduzcan de manera significativa la confianza en una estimación grande del efecto.

La evaluación resumen del riesgo de sesgo se puede considerar en cuatro niveles:

- Resumen del riesgo de sesgo de un estudio a través de los desenlaces: Algunos dominios afectan el riesgo de sesgo entre los desenlaces de un estudio, p.ej. la generación de la secuencia y la ocultación de la secuencia de asignación. Otros dominios como el cegamiento y los datos de desenlace incompletos pueden tener diferentes riesgos de sesgo para diferentes desenlaces dentro de un estudio. Por lo tanto, los revisores no deberían suponer que el riesgo de sesgo es el mismo para todos los desenlaces de un estudio. Además, en general una evaluación resumen del riesgo de sesgo para todos los desenlaces de un estudio despierta poco interés.

- Resumen del riesgo de sesgo de un desenlace dentro de un estudio (entre los dominios): Éste es el nivel recomendado para resumir el riesgo de sesgo de un estudio, ya que algunos riesgos de sesgo pueden ser diferentes para diferentes desenlaces. Una evaluación resumen del riesgo de sesgo para un desenlace debería incluir todos los ítems relevantes para ese desenlace, es decir, los ítems a nivel de estudio como la ocultación de la secuencia de asignación, y los ítems específicos para los desenlaces, como el cegamiento.

- Resumen del riesgo de sesgo de un desenlace a través de los estudios (p.ej. para un metanálisis): Éstas son las principales evaluaciones resumen que harán los revisores e incorporarán en las evaluaciones acerca de la “calidad de la evidencia” en las tablas “Resumen de los hallazgos”, como se describe en el Capítulo 11 (Sección 11.5).

- Resumen del riesgo de sesgo para una revisión en general (de estudios y desenlaces): Podría ser tentador resumir el riesgo general de sesgo de una revisión, pero se debería evitar por dos motivos. Primero, esto requiere juicios de valor acerca de qué desenlaces son fundamentales para una decisión, por lo que éstos se deberían incluir en esta evaluación. Con frecuencia no hay datos disponibles de los estudios incluidos en una revisión para algunos desenlaces que pueden ser fundamentales, como los efectos adversos, y en muy

pocas ocasiones el riesgo de sesgo es el mismo para todos los desenlaces fundamentales para dicha evaluación. En segundo lugar, las evaluaciones acerca de qué desenlaces son fundamentales para una decisión pueden variar de ámbito a ámbito debido a diferencias en los valores y a diferencias en otros factores como el riesgo inicial. Por lo tanto, las evaluaciones acerca de la evidencia de riesgo de sesgo general entre los estudios y los desenlaces se deberían realizar en un contexto específico, por ejemplo, en el contexto de las guías de práctica clínica y no en el contexto de revisiones sistemáticas que tienen como objetivo informar decisiones en una variedad de ámbitos.

Los revisores deberían hacer evaluaciones explicitas acerca del riesgo de sesgo para los desenlaces importantes dentro y entre los estudios. Esto requiere identificar los dominios más importantes ("dominios clave") que se incluyen en estas evaluaciones resumen. La Tabla 8.7.a proporciona un posible enfoque para hacer evaluaciones resumen del riesgo de sesgo para desenlaces importantes dentro y entre los estudios.

Tabla 8.7.a: Posible enfoque para las *evaluaciones resumen* del riesgo de sesgo para cada desenlace importante (entre los dominios) dentro y entre los estudios

Riesgo de sesgo	Interpretación	Dentro de un estudio	Entre los estudios
Bajo riesgo de sesgo.	Sesgo plausible, poco probable que altere significativamente los resultados.	Bajo riesgo de sesgo para todos los dominios clave.	La mayor parte de la información proviene de estudios con bajo riesgo de sesgo
Riesgo de sesgo poco claro.	Sesgo plausible que hace surgir algunas dudas acerca de los resultados.	Riesgo de sesgo poco claro para uno o más dominios clave.	La mayor parte de la información proviene de estudios con riesgo de sesgo bajo o poco claro.
Alto riesgo de sesgo.	Sesgo plausible que debilita seriamente la confianza en los resultados.	Alto riesgo de sesgo para uno o más dominios clave.	La proporción de la información que proviene de estudios con alto riesgo de sesgo es suficiente para afectar la interpretación de los resultados.

8.8 Incorporación de las evaluaciones en los análisis

8.8.1 Introducción

A menudo las consideraciones estadísticas incluyen un balance entre sesgo y precisión. Un metanálisis que incluya todos los estudios elegibles puede producir un resultado con gran precisión (un intervalo de confianza estrecho), pero estar seriamente sesgado debido a deficiencias en la realización de alguno de los estudios. Por otra parte, si se incluyen solamente los estudios con bajo riesgo de sesgo en todos los dominios evaluados se puede producir un resultado no sesgado pero impreciso (si sólo hay pocos estudios de alta calidad).

Cuando se realizan y presentan los metanálisis los revisores deberían abordar los riesgos de sesgo en los resultados de los estudios incluidos. No es apropiado presentar los análisis e interpretaciones basados en todos los estudios, e ignorar las deficiencias identificadas durante la evaluación del riesgo de sesgo. Mientras mayor sea la proporción de estudios evaluados como con alto riesgo de sesgo, más precaución se debería tener al analizar e interpretar sus resultados.

8.8.2 Exploración de la repercusión del riesgo de sesgo

8.8.2.1 Gráficos de los resultados según el riesgo de sesgo

En la discusión que sigue se hará referencia a las comparaciones de los resultados según los dominios de sesgo individuales. Sin embargo, dichas comparaciones también se pueden realizar según el riesgo de sesgo resumido a nivel del estudio (ver Sección 8.7).

Los gráficos de la estimación del efecto de la intervención (p.ej. gráficos de bosque) estratificados según el riesgo de sesgo pueden ser una manera útil de comenzar a examinar la posibilidad de que el sesgo afecte los resultados del metanálisis. Los gráficos de bosque ordenados según las evaluaciones de cada ítem del "Riesgo de sesgo" están disponibles en RevMan 5. Dichos gráficos proporcionan una impresión visual de las contribuciones relativas de los estudios con riesgo bajo, poco claro o alto de sesgo, así como del grado de las diferencias en las estimaciones del efecto de la intervención entre los estudios con riesgo de sesgo bajo, poco claro o alto. Habitualmente será apropiado limitar dichos gráficos de bosque a los dominios de sesgo claves (ver Sección 8.7).

8.8.2.2 Estudios evaluados como riesgo poco claro de sesgo

Los estudios se evalúan como riesgo poco claro de sesgo cuando hay muy pocos detalles disponibles para evaluarlos como de "alto" o "bajo" riesgo; cuando el riesgo de sesgo verdaderamente se desconoce a pesar de que exista información suficiente acerca de la realización, o cuando un ítem no es relevante para un estudio (por ejemplo, debido a que el estudio no menciona ninguno de los desenlaces en el grupo de desenlaces a los cuales se aplica el ítem). Cuando el primer motivo es el principal, es razonable suponer que el sesgo promedio en los resultados de dichos estudios evaluados será menor que en los estudios con alto riesgo de sesgo, ya que cuando se realizaron algunos estudios evaluados como "poco claro", de hecho se habría evitado el sesgo. La evidencia limitada a partir de estudios empíricos que examinaron las categorías "alto" e "poco claro" por separado lo confirma. Por ejemplo, el estudio de Schulz y cols. encontró que los *odds ratios* para la intervención se exageraron en un 41% en los ensayos con ocultación inadecuada (alto riesgo de sesgo) y en un 30% en los ensayos con ocultación poco clara (riesgo poco claro de sesgo) (Schulz 1995b). Sin embargo, la mayoría de los estudios empíricos han combinado las categorías "alto" e "poco claro", que se compararon luego con la categoría "bajo".

Se recomienda que los revisores no combinen estudios con riesgo de sesgo "bajo" e "poco claro" en los análisis, a menos que proporcionen motivos específicos para creer que es probable que estos estudios se hayan realizado de forma que se evitó el sesgo. En el resto de esta sección se asumirá que los estudios evaluados como bajo riesgo de sesgo se tratarán como una categoría separada.

8.8.2.3 Metaregresión y comparaciones de subgrupos

Las comparaciones formales de los efectos de la intervención según el riesgo de sesgo se pueden realizar mediante la metaregresión (ver Capítulo 9, Sección 9.6.4). Para los estudios con desenlaces dicotómicos, los resultados de los análisis de metaregresión son más útiles si se expresan como cocientes de *odds ratios* (o cocientes de riesgos) que comparan los resultados de los estudios con riesgo de sesgo alto o poco claro con los de los estudios con bajo riesgo de sesgo.

$$\text{Cociente de Odds Ratios} = \frac{\text{Odds ratio de intervención en estudios de alto o incierto riesgo de sesgo}}{\text{Odds ratio de intervención en estudios de bajo riesgo de sesgo}}$$

Alternativamente, se pueden realizar comparaciones separadas de alto versus bajo y poco claro versus bajo. Para los estudios con resultados continuos (p.ej. presión sanguínea) los

efectos de la intervención se expresan como diferencias de medias entre los grupos de intervención, y los resultados de los análisis de metaregresión corresponden a las diferencias de las diferencias de medias.

Si la estimación del efecto de la intervención es el mismo en los estudios con riesgo de sesgo alto y poco claro que en los estudios con bajo riesgo de sesgo entonces el cociente de los *odds ratios* (o cocientes de riesgo) será igual a 1, mientras que la diferencia entre las diferencias de medias será igual a cero. Como se explica en la Sección 8.2.3, la evidencia empírica a partir de la obtención de los metanálisis incluidos en los estudios metaepidemiológicos indican que, como promedio, las estimaciones del efecto de la intervención tienden a ser más exageradas en los estudios con un riesgo de sesgo alto o poco claro que en los estudios con bajo riesgo de sesgo.

Cuando un metanálisis incluye varios estudios, los análisis de metaregresión pueden incluir más de un dominio (p.ej. ocultamiento de la asignación y cegamiento).

Los resultados de los análisis de metaregresión incluyen un intervalo de confianza para el cociente de los *odds ratios*, y un valor de p para la hipótesis nula de que no existe diferencia entre los resultados de los estudios con riesgo de sesgo alto o poco claro y bajo. Debido a que habitualmente los metanálisis contienen un escaso número de estudios, el cociente de los *odds ratios* con frecuencia se calcula de forma poco precisa. Por lo tanto, es importante no establecer conclusiones, según un valor no significativo de p, de que no hay diferencias entre los resultados de los estudios con riesgo de sesgo alto o poco claro y bajo, y por este motivo, que el sesgo no repercutió sobre los resultados. A menudo el examen del intervalo de confianza mostrará que las diferencias entre los estudios con riesgo de sesgo alto o poco claro y con riesgo de sesgo bajo son consistentes tanto con ningún sesgo como con un efecto significativo de sesgo.

Una prueba para las diferencias entre los subgrupos proporciona una alternativa a la metaregresión para examinar un ítem único (p.ej. comparar estudios con ocultación adecuada frente a inadecuada de la asignación). Dentro del marco de un metanálisis de efectos fijos, estas pruebas están disponibles en RevMan 5. Sin embargo, estos valores de p son de uso limitado sin sus correspondientes intervalos de confianza, y en todo caso serán demasiado pequeños si existe heterogeneidad, ya sea dentro o entre los subgrupos.

8.8.3 Inclusión de las evaluaciones del “riesgo de sesgo” en los análisis

En general, los estudios con riesgo de sesgo alto o poco claro deberían tener una ponderación reducida en los metanálisis, comparados con los estudios con bajo riesgo de sesgo (Spiegelhalter 2003). Sin embargo, los métodos estadísticos formales para combinar los resultados de los estudios con riesgo alto y bajo de sesgo no están suficientemente bien desarrollados como para poder recomendar actualmente su uso en las revisiones Cochrane (ver Sección 8.8.4.2). Por lo tanto, el enfoque principal para incorporar las evaluaciones del riesgo de sesgo en las revisiones Cochrane es limitar los metanálisis a los estudios con riesgo de sesgo bajo (o más bajo) o bien estratificar los estudios según su riesgo de sesgo.

8.8.3.1 Posibles estrategias de análisis

Cuando los riesgos de sesgo varían entre los estudios en un metanálisis, hay tres estrategias amplias disponibles para seleccionar qué resultado presentar como el hallazgo principal para un determinado resultado (por ejemplo, para decidir qué resultado presentar en el resumen). La estrategia planificada se debería describir en el protocolo de la revisión.

1. Análisis primario limitado a los estudios con riesgo de sesgo bajo (o bajo y poco claro)

El primer enfoque implica definir un umbral basado en los dominios clave (ver Sección 8.7), de manera que sólo los estudios que cumplan criterios específicos se incluyan en el análisis primario. El umbral se puede determinar mediante los criterios de elegibilidad originales de la

revisión, o mediante un argumento razonable (que puede establecerse a partir de evidencia empírica de sesgo proveniente de estudios meta-epidemiológicos). Si el análisis primario incluye estudios con riesgo poco claro de sesgo, los revisores deberían proporcionar la justificación para esta decisión. De manera ideal, el umbral o el método para determinarlo, se deberían especificar en el protocolo de revisión. Los revisores deberían recordar que todos los umbrales son arbitrarios y que en teoría los estudios pueden caer en cualquier lado del espectro, desde "libre de sesgo" a "indudablemente sesgado". Mientras más alto sea el umbral, más similares serán los estudios en cuanto al riesgo de sesgo, pero puede provocar que haya un menor número de estudios disponibles. Se aconseja a los autores que limiten su primer análisis de esta forma que realicen análisis de sensibilidad para mostrar cómo las conclusiones podrían afectarse si se incluyeran estudios de alto riesgo de sesgo.

2. Presentación de análisis múltiples (estratificados)

Al estratificar por el resumen del riesgo de sesgo puede producir por lo menos tres intervenciones del efecto de la intervención: a partir de los estudios con alto y bajo riesgo de sesgo y a partir de todos los estudios. Es posible presentar dos o más análisis con igual importancia que incorporen diferentes criterios de inclusión, por ejemplo, uno que incluya todos los estudios y uno que incluya solamente los que tienen un bajo riesgo de sesgo. Esto evita la necesidad de tomar decisiones difíciles, pero puede ser confuso para los lectores. En particular, las personas que deben tomar decisiones habitualmente necesitan una estimación única del efecto. Además, las tablas "Resumen de los hallazgos" generalmente sólo presentarán un resultado único para cada resultado. Por otra parte, un gráfico de bosque estratificado presenta toda la información de manera transparente.

La elección entre la primera y la segunda estrategia se debería basar en el contexto de una determinada revisión y el equilibrio entre el potencial para el sesgo y la pérdida de precisión cuando se excluyen estudios de riesgo de sesgo alto y poco claro. Como se explica en la Sección 8.8.2.3, la falta de diferencias estadísticamente significativas entre los estudios de alto y bajo riesgo de sesgo no deberían interpretarse como ausencia de sesgo, porque los análisis de metaregresión tienen poco poder.

3. Presentación de todos los estudios proporcionando una discusión narrativa del riesgo de sesgo

El enfoque más sencillo para incorporar las evaluaciones del riesgo de sesgo es presentar una estimación del efecto de la intervención basada en todos los estudios disponibles, junto con una descripción del riesgo de sesgo en los dominios individuales, o una descripción del riesgo de sesgo resumen entre los estudios. Esta es la única opción factible cuando todos los estudios tienen un alto riesgo, todos tienen un riesgo incierto o todos tienen un bajo riesgo de sesgo. Sin embargo, cuando los estudios tienen riesgos de sesgo diferentes, no se recomienda este enfoque por dos motivos. Primero, las descripciones detalladas del riesgo de sesgo en la sección "Resultados", junto con una interpretación cuidadosa en la sección "Discusión", a menudo se perderán en las "Conclusiones", el "Resumen" y el "Resumen de los hallazgos", por lo que la interpretación final ignorará el riesgo de sesgo. En segundo lugar, con dicho análisis no es posible reducir la ponderación de los estudios con alto riesgo de sesgo, lo que dará lugar a una intervención que en general es demasiado precisa y al mismo tiempo está potencialmente sesgada.

Cuando el análisis primario se basa en todos los estudios, las evaluaciones resumen del riesgo de sesgo deberían incorporarse a medidas explícitas de la calidad de la evidencia para cada resultado importante, por ejemplo utilizando el sistema GRADE (Guyatt 2008). Esto puede ayudar a asegurar que las evaluaciones sobre el riesgo de sesgo y otros factores que afecten la calidad de la evidencia, como la imprecisión, la heterogeneidad y el sesgo de publicación se han tenido en cuenta apropiadamente al interpretar los resultados de la revisión (Ver Capítulo 11, Sección 11.5 y Capítulo 12, Sección 12.2).

8.8.4 Otros métodos para tratar con el riesgo de sesgo

8.8.4.1 Ponderación directa

Se han descrito métodos para ponderar los estudios en el metanálisis según su validez o el riesgo de sesgo (Detsky 1992). El método estadístico habitual para combinar los resultados de múltiples estudios es ponderar los estudios según la cantidad de información que proporcionen (más específicamente, mediante las varianzas inversas de sus estimaciones del efecto). Esto da a los estudios con resultados más precisos (intervalos de confianza más estrechos) una ponderación mayor. También es posible ponderar los estudios de forma adicional según la validez, de manera que los estudios más válidos tengan una mayor influencia sobre el resultado resumen. Se puede utilizar una combinación de las varianzas inversas y las evaluaciones de la validez. La objeción principal a este enfoque es que requiere un resumen numérico de la validez de cada estudio y que no hay evidencia empírica para determinar cuánta ponderación asignarle a los diferentes dominios de sesgo. Además, el promedio ponderado resultante estará sesgado si alguno de los estudios está sesgado. Se debería evitar la ponderación directa de las estimaciones del efecto según la validez o las evaluaciones del riesgo de sesgo (Greenland 2001).

8.8.4.2 Enfoques bayesianos

Los análisis bayesianos permiten la incorporación de la información o la opinión externa en la naturaleza del sesgo (ver Capítulo 16, Sección 16.8) (Turner 2008). Las distribuciones previas para sesgos específicos en las estimaciones del efecto de la intervención se pueden basar en evidencia empde sesgo, en opiniones de expertos obtenidas previamente o en argumentos razonados. Los métodos bayesianos de ajuste de los metanálisis para los sesgos son actualmente tema de investigación y no están suficientemente desarrollados para adoptarlos ampliamente.

8.9 Generación de la secuencia

8.9.1 Justificación para las inquietudes acerca del sesgo

Bajo el dominio de la generación de la secuencia en la herramienta de la Colaboración para evaluar el riesgo de sesgo, se menciona si el estudio utilizó o no una secuencia de asignación aleatoria. Este es el primero de dos dominios en la herramienta de la Colaboración que aborda el proceso de asignación; el segundo es la ocultación de la secuencia de asignación (ocultación de la asignación). En primer lugar se explican las diferencias entre estos dominios.

El punto de partida para un estudio de intervención no sesgado es el uso de un mecanismo que asegure que el mismo tipo de participantes reciba cada intervención. Es necesario considerar numerosos procesos interrelacionados. Primero, se debería utilizar una secuencia de asignación que, si se implementa de manera perfecta, equilibraría los factores pronósticos, como promedio, de manera similar entre los grupos de intervención. La asignación aleatoria desempeña un papel fundamental aquí. Es posible argumentar que otras reglas de asignación como la alternancia (alternar entre dos intervenciones) o la rotación (asignar cíclicamente más de dos intervenciones), pueden lograr el mismo efecto (Hill 1990). Sin embargo, una regla teóricamente no sesgada no es suficiente para evitar el sesgo en la práctica. Si es posible prever las asignaciones futuras al predecirlas o conocerlas, entonces puede surgir el sesgo de selección debido al reclutamiento selectivo y al no reclutamiento de participantes en un estudio según la siguiente asignación a la intervención.

Es posible anticipar las asignaciones futuras por varios motivos. Estos incluyen (i) conocimiento de una regla de asignación determinista como la alternancia, la fecha de nacimiento o el día de ingreso; (ii) conocimiento de la secuencia de las asignaciones, sean

aleatorias o no (p.ej. si una secuencia de asignaciones aleatorias se coloca en la pared); (iii) capacidad de predecir las asignaciones de forma exitosa, según las asignaciones previas (que en algunas ocasiones es posible cuando los métodos de asignación aleatorios que se utilizan intentan asegurar una proporción exacta de asignaciones a diferentes intervenciones). Las interrelaciones complejas entre los aspectos teóricos y prácticos de la asignación en los estudios de intervención hacen que la evaluación del sesgo de selección sea un reto. Quizás el aspecto práctico más importante sea la ocultación de la secuencia de asignación, es decir, el uso de mecanismos para evitar el conocimiento previo de la siguiente asignación. Este aspecto se ha evaluado históricamente en las revisiones Cochrane, con una justificación empírica. La ocultación de la secuencia de asignación se menciona en la herramienta como un dominio separado (ver Sección 8.10).

La asignación aleatoria permite que la secuencia sea impredecible. Una secuencia impredecible, combinada con la ocultación de la secuencia de asignación, debería ser suficiente para evitar el sesgo de selección. Sin embargo, el sesgo de selección puede surgir a pesar de la asignación aleatoria si las asignaciones aleatorias no se ocultan y el sesgo de selección puede surgir (en teoría al menos) a pesar de la ocultación de la secuencia de asignación si la secuencia subyacente no es aleatoria. Se reconoce que una secuencia de asignación no siempre es completamente impredecible, incluso si se utilizan mecanismos para la ocultación de la asignación. Este puede ser algunas veces el caso, por ejemplo, cuando se utiliza la asignación en bloques, y se conocen todas las asignaciones después del reclutamiento. No obstante, esta situación especial no se considera bajo el ítem de la generación de la secuencia ni el ocultamiento de la asignación, y se aborda como un aspecto separado en la Sección 8.14.1.4.

Los estudios metodológicos han evaluado la importancia de la generación de la secuencia. Al menos cuatro de estos estudios han evitado los factores de confusión debidos a la enfermedad o la intervención, lo cual es fundamental para la evaluación (Schulz 1995b, Moher 1998, Kjaergard 2001, Siersma 2007). Se ha observado que la generación inadecuada de las secuencias de asignación se asocia con efectos de la intervención sesgados entre los estudios (Als-Nielsen 2004). En un estudio que limitó el análisis a 79 ensayos que habían descrito una ocultación adecuada de la secuencia de asignación, los ensayos con una generación inadecuada de la secuencia como promedio produjeron estimaciones exageradas de los efectos de la intervención, comparados con los ensayos con una generación adecuada de la secuencia (*odds ratio* relativo 0,75; IC del 95%: 0,55 a 1,02; p = 0,07). Estos resultados indican que si las asignaciones no son aleatorias, es en cierta medida posible descifrar la secuencia, incluso con una ocultación aparentemente adecuada de la secuencia de asignación (Schulz 1995b).

8.9.2 Evaluación del riesgo de sesgo con relación a la generación adecuada o inadecuada de la secuencia

A menudo la generación de la secuencia se aborda de forma inapropiada en las fases de diseño e implementación de los estudios controlados aleatorizados (ECA), y con frecuencia se obvia en las publicaciones, lo que provoca problemas importantes al evaluar el riesgo de sesgo. Las siguientes consideraciones pueden ayudar a los revisores a evaluar si la generación de la secuencia es adecuada para proteger contra el sesgo cuando se utiliza la herramienta de la Colaboración (Sección 8.5).

8.9.2.1 Métodos adecuados de generación de la secuencia

El uso de un componente aleatorio debería ser suficiente para una generación adecuada de la secuencia.

A la asignación aleatoria sin limitaciones para generar una secuencia de asignación se le llama asignación aleatoria simple o asignación aleatoria sin limitaciones. En principio, la misma se puede lograr al asignar las intervenciones mediante métodos como el lanzamiento

repetido de la moneda, el lanzamiento de los dados o el barajar cartas previamente mezcladas (Schulz 2002c, Schulz 2006). Es más habitual lograrla mediante la referencia a una lista de números aleatorios publicada, o a una lista de asignaciones aleatorias generadas por ordenador. En los ensayos que utilizan muestras grandes (generalmente al menos 100 sujetos en cada grupo asignado aleatoriamente) (Schulz 2002c, Schulz 2002d, Schulz 2006), la asignación aleatoria simple genera grupos de comparación con tamaños de muestra relativamente similares. En los ensayos que utilizan muestras pequeñas, en algunas ocasiones la asignación aleatoria simple dará lugar a una secuencia de asignación en la que los grupos difieren, debido al azar, de forma bastante significativa en el tamaño o la aparición de factores pronósticos (es decir, variación de "mezcla de casos") (Altman 1999).

Ejemplo (de bajo riesgo de sesgo): Los dos grupos de comparación se generaron mediante la asignación aleatoria simple, con una proporción de asignación similar, al utilizar una tabla de números aleatorios.

Algunas veces se utiliza la asignación aleatoria limitada para generar una secuencia que asegure proporciones de asignación específicas a los grupos de intervención (p.ej. 1:1). La asignación en bloques (bloques aleatorios permutados) es una forma frecuente de asignación aleatoria limitada (Schulz 2002c, Schulz 2006). Los bloques aseguran que el número de participantes a asignar a cada grupo de comparación estará equilibrado entre bloques de, por ejemplo, cinco en un grupo y cinco en el otro por cada diez participantes ingresados de forma consecutiva. El tamaño del bloque puede variar de forma aleatoria para reducir la probabilidad de conocimiento previo de la asignación a la intervención.

Ejemplo (de bajo riesgo de sesgo): Se utilizó la asignación aleatoria en bloques para formar la lista de asignación para los dos grupos de comparación. Se utilizó un generador de números aleatorios por ordenador para seleccionar los bloques aleatorios permutados con un tamaño de bloque de ocho y una proporción de asignación similar.

También es frecuente la asignación alatoria estratificada, en la cual la asignación aleatoria limitada se realiza de forma separada dentro del estrato. Esto genera esquemas de asignación aleatoria separados para subgrupos de participantes definidos según factores pronósticos potencialmente importantes como la gravedad de la enfermedad y los centros de estudio. Si se utiliza una asignación aleatoria simple (en lugar de limitada) en cada estrato, entonces la estratificación no tendría efecto pero la asignación aleatoria aún sería válida. El riesgo de sesgo se puede evaluar de la misma manera, independientemente de si el ensayo declara o no ser estratificado.

Otro enfoque que incorpora los conceptos generales de estratificación y asignación aleatoria limitada es la minimización, que se puede utilizar para formar grupos pequeños muy similares con respecto a numerosas características. No se debería considerar automáticamente que el uso de la minimización implique un riesgo de sesgo para un estudio. Sin embargo, algunos metodólogos son cautelosos acerca de la aceptabilidad de la minimización, especialmente cuando se utiliza sin un componente aleatorio, mientras que otros la consideran muy atractiva (Brown 2005).

Otros tipos adecuados de asignación al azar que se utilizan algunas veces son la asignación aleatoria por moneda o urna sesgada, la asignación aleatoria por reemplazo, la asignación aleatoria mixta y la asignación aleatoria máxima (Schulz 2002c, Schulz 2002d, Berger 2003). Si se encuentran estos u otros enfoques, puede ser necesario consultar a un estadístico.

8.9.2.2 Métodos inadecuados de generación de la secuencia

A los métodos sistemáticos como la alternancia, la asignación basada en la fecha de nacimiento, el número de historia clínica y la fecha de presentación, se les llama algunas veces "cuasialeatorios". La alternancia (o rotación, para más de dos grupos de intervención)

podría dar lugar en principio a grupos similares, pero no así muchos otros métodos sistemáticos de generación de la secuencia. Por ejemplo, el día en el cual el paciente ingresa en el hospital no depende solamente del azar.

Una debilidad importante de todos los métodos sistemáticos es que la ocultación del esquema de asignación habitualmente es imposible, lo que permite el conocimiento previo de la asignación de la intervención entre los que reclutan a los participantes en el estudio, así como la asignación sesgada (ver Sección 8.10).

Ejemplo (de alto riesgo de sesgo): Los pacientes se asignaron al grupo de intervención según la semana del mes.

Ejemplo (de alto riesgo de sesgo): Los pacientes nacidos en días pares se asignaron al Tratamiento A y los pacientes nacidos en días impares se asignaron el Tratamiento B.

8.9.2.3 Métodos de generación de la secuencia con riesgo poco claro de sesgo

Con frecuencia el planteamiento sencillo “se asignaron de forma aleatorizada” o “se utilizó un diseño aleatorizado” no es suficiente para garantizar que la secuencia de asignación fue verdaderamente aleatoria. Los autores con alguna frecuencia utilizan el término “aleatorizado” incluso cuando no se justifica; muchos ensayos con una asignación sistemática declarada son descritos por los autores como aleatorizados. Si existen dudas, entonces se debería considerar que la generación de la secuencia fue poco clara.

En algunas ocasiones los autores del ensayo proporcionan alguna información, pero definen de forma incompleta su enfoque y no confirman un componente aleatorio en el proceso. Por ejemplo, los autores pueden señalar que se utilizó una asignación aleatoria en bloques, pero no se especificó el proceso de selección de los bloques, como una tabla de números aleatorios o un generador de números aleatorios por ordenador. En este caso la generación de la secuencia también se debería clasificar como poco clara.

8.10 Ocultación de la secuencia de asignación

8.10.1 Justificación para las inquietudes acerca del sesgo

La generación aleatoria de la secuencia es necesaria pero no es una protección suficiente contra el sesgo en la asignación de la intervención. Es probable que los esfuerzos realizados para generar secuencias no sesgadas e impredecibles no sean efectivos si las secuencias no están protegidas mediante una ocultación adecuada de la secuencia de asignación para las personas involucradas en el reclutamiento y la asignación de los participantes.

El conocimiento de la siguiente asignación (por ejemplo, a partir de una tabla de números aleatorios abiertamente colocada en una pizarra de anuncios) puede dar lugar al reclutamiento selectivo de los participantes basado en los factores pronósticos. Los participantes que se habrían asignado a una intervención considerada “inapropiada” pueden ser rechazados. Otros participantes pueden ser dirigidos deliberadamente a la intervención “apropiada”, lo cual a menudo se puede acompañar de retraso en el ingreso de un participante en el ensayo hasta que aparezca la siguiente asignación apropiada. Puede ocurrir que se descifren los esquemas de asignación incluso si se intentó la ocultación. Por ejemplo, es posible abrir los sobres de asignación que no estén cerrados, mientras que los sobres traslúcidos se pueden colocar contra una luz brillante para revelar su contenido (Schulz 1995a, Schulz 1995b, Jüni 2001). Informes personales indican que los investigadores han descifrado muchos esquemas de asignación debido a que los métodos de ocultación fueron inadecuados (Schulz 1995a).

Evitar estos sesgos de selección depende de evitar el conocimiento previo de la asignación a la intervención. Las decisiones sobre la elegibilidad de los participantes y su decisión de dar el consentimiento informado se deberían tomar sin conocer la siguiente asignación. La adecuada ocultación de la secuencia de asignación evita que aquellos que ingresan a los participantes en un estudio conozcan la próxima asignación.

Algunos estudios metodológicos han examinado si la ocultación de la secuencia de asignación se asocia a la magnitud de las estimaciones del efecto en los ensayos clínicos controlados, a la vez que evita los factores de confusión debidos a la enfermedad o la intervención. Un análisis combinado de siete estudios metodológicos encontró que las estimaciones del efecto de los ensayos con ocultación inadecuada de la asignación o descripción poco clara de la técnica utilizada para la ocultación de la asignación fueron como promedio 18% más "beneficiosas" que las estimaciones del efecto de los ensayos con ocultación adecuada de la asignación (intervalo de confianza del 95%: 5% a 29%) (Pidal 2007). Un metanálisis detallado reciente de tres de estos grupos de datos combinados (1346 ensayos de 146 metanálisis) arrojó alguna luz sobre la heterogeneidad de estos estudios. Las estimaciones del efecto de la intervención se exageraron cuando hubo ocultación inadecuada de la asignación en los ensayos en los que se analizó un resultado subjetivo, pero hubo poca evidencia de sesgo en los ensayos con desenlaces objetivos (Wood 2008).

8.10.2 Evaluación del riesgo de sesgo con relación a la ocultación adecuada o inadecuada de la asignación de la secuencia

Las siguientes consideraciones pueden ayudar a los revisores a evaluar si la ocultación de la asignación fue suficiente para proteger contra el sesgo, cuando utilizan la herramienta de la Colaboración (ver Sección 8.5).

La ocultación adecuada de la secuencia de asignación asegura la implementación estricta de una secuencia de asignación sin el conocimiento de las asignaciones a la intervención. Los métodos para la ocultación de la asignación se refieren a las técnicas utilizadas para implementar la secuencia, no para generarla (Shulz 1995b). Sin embargo, la mayoría de las *secuencias* de asignación que se consideran inadecuadas, como la asignación según el día de ingreso o el número de historia clínica, no se pueden ocultar de forma adecuada, por lo que no cumplen ambos ítems. Es teóricamente posible, aunque poco probable, que una secuencia inadecuada se oculte de forma adecuada (la persona responsable del reclutamiento y la asignación a las intervenciones no estaría al tanto de que la secuencia que se implementa no es apropiada). Sin embargo, es frecuente que una secuencia de asignación adecuada (aleatoria) no se haya ocultado de forma adecuada, por ejemplo, si la secuencia se coloca en la pared del salón del personal.

Algunos revisores confunden la ocultación de la asignación con el cegamiento de las intervenciones asignadas. La ocultación de la asignación intenta evitar el sesgo de selección en la asignación a la intervención al proteger la secuencia de la asignación *antes y hasta* la asignación, y siempre se puede implementar de forma exitosa independientemente del tema de estudio (Schulz 1995b, Jüni 2001). Por el contrario, el cegamiento intenta evitar el sesgo de realización y detección al proteger la secuencia *después* de la asignación (Jüni 2001, Schulz 2002a), y no siempre es posible implementarla, por ejemplo, en los ensayos que comparan tratamientos médicos y quirúrgicos. Por lo tanto, la ocultación de la asignación hasta el momento de la asignación a la intervención y el cegamiento después de este momento se refieren a diferentes fuentes de sesgo y difieren en cuanto a su factibilidad.

La importancia de la ocultación de la asignación puede depender del grado en el cual los participantes potenciales en el estudio tienen diferentes pronósticos, si existen creencias fuertes entre los investigadores y los participantes con respecto a los beneficios o daños de las intervenciones asignadas y si todas las personas involucradas aceptan las incertidumbres acerca de las intervenciones (Schulz 1995a). Entre los diferentes métodos utilizados para ocultar la asignación, la asignación aleatoria central por un tercero es quizás la más

conveniente. Los métodos que utilizan sobres son más susceptibles a la manipulación que otros enfoques (Schulz 1995b). Si los investigadores utilizan sobres, deberían desarrollar y monitorizar el proceso de asignación para mantener la ocultación. Además de utilizar sobres cerrados, oscuros y numerados de forma secuencial, deberían asegurar que los sobres se abran secuencialmente y sólo después de que el sobre se haya asignado de forma irreversible al participante.

8.10.2.1 Métodos adecuados de ocultación de la secuencia de asignación

La Tabla 8.10.a proporciona los criterios mínimos para una evaluación de la ocultación adecuada de la secuencia de asignación (izquierda), así como los criterios ampliados que proporcionan una confirmación adicional de que la ocultación de la secuencia de asignación fue verdaderamente adecuada (derecha).

Ejemplos (de bajo riesgo de sesgo) [descripciones publicadas de procedimientos de ocultación que se consideraron adecuados, compilados por Schulz y Grimes (Schulz 2002b)]:

"… que combinó números codificados con la asignación del fármaco. Cada bloque de diez números se transmitió desde la oficina central a una persona que actuó como la autoridad de la asignación aleatoria en cada centro. Este individuo (un farmacéutico o una enfermera no involucrados en la atención a los pacientes del ensayo e independientes del investigador local) fue responsable de la asignación, preparación y recuento de la infusión del ensayo. La infusión del ensayo se preparó en un lugar separado, luego se le llevó a la enfermera de cabecera cada 24 horas. La enfermera la infundió al paciente con el goteo adecuado. Por lo tanto, el esquema de asignación aleatoria se les ocultó a todos los proveedores de atención, médicos de la sala y otro personal de investigación." (Bellomo 2000).

"… se ocultó en sobres oscuros, cerrados y numerados de forma secuencial, y la guardó el farmacéutico del hospital en los dos centros."(Smilde 2001).

"Los tratamientos se asignaron de forma centralizada después de la verificación telefónica de lo adecuado de los criterios de inclusión …" (de Gaetano 2001).

"El Departamento de Farmacia del Glenfield Hospital realizó la asignación aleatoria, distribuyó los agentes del estudio y guardó los códigos del ensayo, que se revelaron después del estudio." (Brightling 2000).

Tabla 8.10.a: Criterios mínimos y ampliados para evaluar la ocultación de la secuencia de asignación como adecuada (bajo riesgo de sesgo)

Criterios mínimos para una evaluación de de la ocultación adecuada de la secuencia de asignación	Criterios ampliados que proporcionan confirmación adicional
Asignación aleatoria central.	La oficina central de asignación aleatoria está lejos de los centros de reclutamiento de pacientes. Los detalles del participante se proporcionaron, por ejemplo, por teléfono, fax o correo electrónico y la secuencia de asignación se ocultó a los individuos del personal de la oficina de asignación al azar hasta que el participante se registró de forma irreversible.
Envases de fármacos numerados de forma secuencial.	Los envases de fármacos preparados por una farmacia independiente se numeraron y se abrieron de forma secuencial. Los envases tenían una apariencia y peso idénticos y eran a prueba de trampas.
Sobres cerrados, oscuros y numerados de forma secuencial.	Los sobres se numeraron y abrieron de forma secuencial solamente después de que los detalles del participante se escribieron en el sobre. Mediante un papel sensible a la presión o un papel carbón colocado dentro del sobre se transfirieron los detalles del participante a la tarjeta de asignación. Una tarjeta de cartón o una hoja de aluminio dentro del sobre mantuvieron el sobre impermeable a la luz intensa. Los sobres se cerraron mediante cinta de seguridad a prueba de trampas.

8.11 Cegamiento de los participantes y de personal

8.11.1 Justificación para las inquietudes acerca del sesgo

En un ensayo clínico es posible cegar a diferentes tipos de participantes y de personal: ver Tabla 8.11.a. Los dos primeros dominios de la herramienta que específicamente abordan el cegamiento se focalizan en los participantes y el personal (los proveedores de la atención sanitaria). Su falta de cegamiento podría sesgar los resultados, al afectar los resultados de los participantes en el ensayo. Esto puede ser debido a una falta de expectativas en el grupo control o bien a diferentes conductas en los grupos de intervención (por ejemplo, diferentes pérdidas, distintos cruzamientos con una intervención alternativa o una administración distinta de cointervenciones).

No se dispone de evidencia empírica del sesgo en relación con la falta de cegamiento de los participantes y del personal. No obstante, hay evidencia para los estudios descritos como ‘ciego’ o ‘doble-ciego’, que habitualmente incluye el cegamiento de uno o varios de este grup de personas. En estudios empíricos, se ha mostrado que la falta de cegamiento en ensayos aleatorizados se asocia con un una estimación del efecto de la intervención más exagerada, como promedio en un 9%, medida como *odds ratios* (Pildal 2007). Estos estudios han analizado varios resultados, algunos de los cuales son objetivos. Se ha observado que la estimación del efecto está más sesgada, como promedio, en los ensayos con resultados más subjetivos (Wood 2008). La falta de cegamiento también pudiera provocar sesgo debido a investigaciones adicionales o cointervenciones, independientemente del tipo de resultado, si las mismas ocurren de forma diferencial entre los grupos de intervención.

El cegamiento puede ser imposible para al menos algunas personas (p.ej. la mayoría de los pacientes a los que se les realiza una cirugía). Sin embargo, dichos estudios pueden utilizar otras medidas para reducir el riesgo de sesgo, como tratar a los pacientes según un protocolo estricto para reducir el riesgo de conductas diferenciales por los pacientes y los proveedores de atención sanitaria. El intento de cegar a los participantes y al personal no asegura un cegamiento exitoso en la práctica. El cegamiento puede estar comprometido para la mayoría de las intervenciones. Para muchos ensayos farmacológicos cegados, los efectos secundarios de los fármacos pueden permitir detectar qué intervención reciben algunos participantes, a menos que el estudio compare dos intervenciones bastante similares, p.ej. fármacos con efectos secundarios similares o utilicen un placebo activo (Boutron 2006).

En los estudios cegados, especialmente los ensayos controlados con placebo, pueden existir inquietudes con respecto a si los participantes del estudio estaban verdaderamente cegados (y algunas veces si también lo estaban quienes atendían a los pacientes). Numerosos grupos han indicado que sería de interés solicitarles a los participantes al final del estudio que adivinen qué tratamiento recibieron (Fergusson 2004, Rees 2005), y se han publicado algunas revisiones de dichos informes (Fergusson 2004, Hróbjartsson 2007). La evidencia de aciertos mayores del 50% parece indicar que el cegamiento puede haberse roto, pero de hecho puede reflejar sencillamente la experiencia del paciente en el ensayo: un buen resultado o un efecto secundario marcado, tenderán a ser atribuidos con mayor frecuencia a un tratamiento activo, y un resultado deficiente al placebo (Sackett 2007). Por lo tanto, es de esperar observar alguna “adivinación” exitosa cuando existen diferencias en la eficacia o los efectos adversos, y ninguna cuando las intervenciones tienen efectos muy similares, incluso cuando se ha preservado el cegamiento. Como consecuencia, los revisores deben considerar cuidadosamente si toman en cuenta los hallazgos de este tipo de ejercicio.

Tabla 8.11.a: Una nota sobre el cegamiento en los ensayos clínicos

En general, el cegamiento (en ocasiones denominado enmascaramiento) se refere al proceso por el cual los participantes, los proveedores sanitarios y los investigadores, incluyendo las personas que evalúan los resultados, no son conscientes de la asignaciones de las intervenciones después de que los participantes se hayan incluido en el estudio. El cegamiento puede reducir el riesgo de que el conocimiento de la intervención que se recibió, más que la intervención en sí, afecte a los resultados y a las valoraciones de los mismos.

Diferentes tipos de personas pueden estar cegadas en un ensayo clínico (Gøtzsche 1996, Heahr 2006):
1. los pacientes (por ejemplo pacientes o personas sanas);
2. los proveedores sanitarios (por ejemplo los profesionales médicos o de enfermería responsables del cuidado de los pacientes);
3. las personas que evalúan los resultados, inclusive las que recopilan los datos primarios (por ejemplo, miembros del equipo responsables de medir y recopilar los datos) y evaluadores secundarios (por ejemplo comités de adjudicación de resultados externos);
4. las que analizan los datos (por ejemplo estadísticos);
5. las que escriben el manuscrito.
Los primeros dos tipos de personas se abordan en la herramienta mediante el ítem ‘Cegamiento de los participantes y del personal’. El tercero se aborda mediante el ítem ‘Cegamiento de la valoración de los resultados’. Los dos últimos no se tratan explícitamente en la herramienta.

8.11.2 Evaluación del riesgo de sesgo con relación al cegamiento adecuado o inadecuado de los participantes y del personal

Las publicaciones de los estudios describen a menudo el cegamiento en términos amplios, como "doble cegamiento". Este término hace imposible conocer quién estaba cegado (Schulz 2002a). Dichos términos también se utilizan de una forma muy inconsistente (Devereaux 2001, Boutron 2005, Haahr 2006), y la frecuencia de la descripción explícita del cegamiento de los participantes y el personal de un estudio aún es bajo incluso en los ensayos publicados en revistas de alto nivel (Montori 2002), a pesar de las recomendaciones explícitas en la Declaración CONSORT (Moher 2001b). Una revisión de los métodos utilizados para el cegamiento destaca la variedad de métodos utilizados en la práctica (Boutron 2006). Las siguientes consideraciones pueden ayudar a los revisores a evaluar si es probable que el cegamiento utilizado en un estudio fue suficiente para protegerlo contra el sesgo, cuando se utiliza la herramienta de la Colaboración (Sección 8.5).

Cuando se considera el riesgo de sesgo debido a la falta de cegamiento es importante analizar específicamente:

1. quién estaba y no estaba cegado;
2. el riesgo de sesgo en los resultados reales debido a la falta de cegamiento durante el estudio (p.ej. debido a cointervenciones o conducta diferencial).

El riesgo de sesgo puede ser alto para algunos resultados y bajo para otros, incluso si las mismas personas no estaban cegadas en el estudio. Por ejemplo, el conocimiento de la intervención asignada puede repercutir en los resultados conductuales (como el número de visitas al consultorio), mientras que no repercuten en los resultados psicológicos o la mortalidad. Por lo tanto, pudiera ser necesario que las evaluaciones del riesgo de sesgo como resultado de la falta de cegamiento se hagan de forma separada para resultados diferentes.

En lugar de evaluar el riesgo de sesgo para cada resultado por separado, a menudo es conveniente agrupar los resultados con riesgo de sesgo similar (ver Sección 8.5). Por ejemplo, puede haber una evaluación del riesgo de sesgo de todos los resultados subjetivos que sea diferente de una evaluación común del cegamiento para todos los resultados objetivos.

8.12 Cegamiento de los evaluadores

8.12.1 Justificación para las inquietudes acerca del sesgo

En un ensayo clínico es posible cegar a diferentes tipos de personas: ver Tabla 8.11.a. Los dos primeros dominios de la herramienta que específicamente abordan el cegamiento se focalizan en los participantes y el personal (proveedores sanitarios). Su falta de cegamiento podría sesgar los resultados al afectar los resultados de los participantes en el ensayo. Esto puede deberse a una falta de expectativas en el grupo control o bien a conductas diferenciales en los grupos de intervención (por ejemplo, pérdidas diferenciales, distintos cruzamientos con una intervención alternativa o una administración distinta de cointervenciones).

No se dispone de evidencia empírica del sesgo en relación con la falta de cegamiento de los participantes y del personal. No obstante, hay evidencia para los estudios descritos como 'ciego' o 'doble-ciego', que habitualmente incluye el cegamiento de uno o varios de este grup de personas. En estudios empíricos, se ha mostrado que la falta de cegamiento en ensayos aleatorizados se asocia con un una estimación del efecto de la intervención más exagerada, como promedio en un 9%, medida como *odds ratios* (Pildal 2007). Estos estudios han analizado varios resultados, algunos de los cuales son objetivos. Se ha observado que la estimación del efecto está más sesgada, como promedio, en los ensayos con resultados más subjetivos (Wood 2008). La falta de cegamiento también pudiera provocar sesgo debido a

investigaciones adicionales o cointervenciones, independientemente del tipo de resultado, si las mismas ocurren de forma diferencial entre los grupos de intervención.

El cegamiento puede ser imposible para al menos algunas personas (p.ej. la mayoría de los pacientes a los que se les realiza una cirugía). Sin embargo, dichos estudios pueden utilizar otras medidas para reducir el riesgo de sesgo, como tratar a los pacientes según un protocolo estricto para reducir el riesgo de conductas diferenciales por los pacientes y los proveedores de atención sanitaria. El intento de cegar a los participantes y al personal no asegura un cegamiento exitoso en la práctica. El cegamiento puede estar comprometido para la mayoría de las intervenciones. Para muchos ensayos farmacológicos cegados, los efectos secundarios de los fármacos pueden permitir detectar qué intervención reciben algunos participantes, a menos que el estudio compare dos intervenciones bastante similares, p.ej. fármacos con efectos secundarios similares o utilicen un placebo activo (Boutron 2006).

En los estudios cegados, especialmente los ensayos controlados con placebo, pueden existir inquietudes con respecto a si los participantes del estudio estaban verdaderamente cegados (y algunas veces si también lo estaban quienes atendían a los pacientes). Numerosos grupos han indicado que sería de interés solicitarles a los participantes al final del estudio que adivinen qué tratamiento recibieron (Fergusson 2004, Rees 2005), y se han publicado algunas revisiones de dichos informes (Fergusson 2004, Hróbjartsson 2007). La evidencia de aciertos mayores del 50% parece indicar que el cegamiento puede haberse roto, pero de hecho puede reflejar sencillamente la experiencia del paciente en el ensayo: un buen resultado o un efecto secundario marcado, tenderán a ser atribuidos con mayor frecuencia a un tratamiento activo, y un resultado deficiente al placebo (Sackett 2007). Por lo tanto, es de esperar observar alguna “adivinación” exitosa cuando existen diferencias en la eficacia o los efectos adversos, y ninguna cuando las intervenciones tienen efectos muy similares, incluso cuando se ha preservado el cegamiento. Como consecuencia, los revisores deberían considerar cuidadosamente si toman en cuenta los hallazgos de este tipo de ejercicio.

8.12.2 Valoración del riesgo de sesgo en relación al cegamiento adecuado o inadecuado de la evaluación de los resultados

Los estudios a menudo describen el cegamiento en términos amplios, como ‘doble ciego’. Este término impide conocer quién estuvo cegado (Schulz 2002a). Además estos términos se utilizan de manera inconsistente (Devereaux 2001, Boutron 2005, Haahr 2006), y la frecuencia de la descripción explícita del cegamiento de los participantes y del personal sigue siendo baja incluso en los ensayos publicado en las revistas de alto impacto (Montori 2002), a pesar de las recomendaciones de la declaración CONSORT (Moher 2001a). Una revisión de los métodos utilizados para el cegamiento destaca la variedad de métodos utilizados en la práctica (Boutron 2006). Las siguientes consideraciones pueden ayudar a los revisores a evaluar si es probable que el cegamiento utilizado en un estudio fue suficiente para protegerlo contra el sesgo, cuando se utiliza la herramienta de la Colaboración (Sección 8.5).

Cuando se considera el riesgo de sesgo debido a la falta de cegamiento de la evaluación de los resultados es importante considerar específicamente:

1. quién evalúa el resultado;
2. el riesgo de sesgo en la evaluación del resultado (teniendo en cuenta cuán subjetivo u objetivo es un resultado).

Los evaluadores de algunos resultados pueden estar cegados, mientras que los evaluadores de otros resultados, no. Por ejemplo, en un ensayo quirúrgico en el que los pacientes conocen su propia intervención, los resultados notificados por el paciente (como por ejemplo, la calidad de vida) se recopilarían bajo el conocimiento de la intervención recibida, mientras que otros resultados, medidos por un clínico independiente podrían ser cegados (como la capacidad fisica). Es más, el riesgo de sesgo puede ser alto para algunos resultados y bajo para otros, incluso si la misma persona no estaba cegada para el estudio. Por ejemplo, el conocimiento de la intervención asignada puede impactar en los resultados notificados por el

paciente (como el nivel de dolor), pero no impactar en otros resultados como la mortalidad. En muchas circunstancias la evaluación de la mortalidad total se puede considerar no sesgada incluso si los evaluadores conocían las asignaciones de las intervenciones. Por lo tanto, la valoración del riesgo de sesgo resultante de la falta de cegamiento puede ser necesario realizarla de forma separada de los otros resultados.

En lugar de valorar el riesgo de sesgo para cada resultado de forma separada, a menudo es conveniente agrupar resultados con riesgo de sesgo similares (ver Sección 8.5) Por ejemplo, puede haber una valoración común del riesgo de sesgo para todos los resultados subjetivos, que es diferente de una valoración común del cegamiento para todos los resultados objetivos.

8.13 Datos de desenlace incompletos

8.13.1 Justificación para las inquietudes acerca del sesgo

Los datos de desenlace faltantes, debido a desgaste (abandono) durante el estudio o a exclusiones del análisis, hacen surgir la posibilidad de que la estimación del efecto observado esté sesgada. Se debería utilizar el término datos de desenlace incompletos para referirse a las deserciones y las exclusiones. Cuando no esté disponible el resultado de un participante individual se le debería llamar "faltante".

El abandono puede ocurrir por los siguientes motivos:
- Abandono de los participantes, o los mismos son retirados del estudio.
- Los participantes no asisten a la cita en la cual se deberían medir los desenlaces.
- Los participantes asisten a la cita pero no proporcionan datos relevantes.
- Los participantes no pueden completar los diarios o cuestionarios.
- No es posible localizar a los participantes (pérdidas durante el seguimiento).
- Los investigadores del estudio deciden, habitualmente de forma inapropiada, interrumpir el seguimiento.
- Los datos o los registros se perdieron, o no están disponibles por otros motivos.

Además, algunos participantes pueden excluirse de los análisis por los siguientes motivos:
- Posteriormente se encontró que algunos participantes reclutados no eran elegibles.
- Se realizó un análisis "como se trató" (o por protocolo) (en el cual los participantes se incluyen solamente si recibieron la intervención planificada según el protocolo; ver Sección 8.12.2).
- El análisis del estudio excluyó algunos participantes por otros motivos.

Algunas exclusiones de participantes parecen ser justificables, y en ese caso no es necesario considerar que dieron lugar a datos de desenlace faltantes (Fergusson 2002). Por ejemplo, los participantes que se asignaron aleatoriamente pero que posteriormente se encontró que no eran elegibles para el ensayo se pueden excluir, siempre que el descubrimiento de la falta de elegibilidad no esté afectado por la intervención asignada aleatoriamente, y preferiblemente basado en decisiones tomadas sin conocimiento de la asignación. La intención de excluir a dichos participantes se debería especificar antes de observar los datos de los desenlaces.

Frecuentemente se recomienda un análisis por intención de tratar (IT) como la forma menos sesgada de realizar la estimación de los efectos de la intervención en ensayos aleatorizados (Newell 1992): ver Capítulo 16 (Sección 16.2). Los principios de los análisis por IT son:
1. mantener a los participantes en los grupos de intervención a los cuales se asignaron aleatoriamente, con independencia de la intervención que realmente recibieron;
2. medir los datos de desenlace en todos los participantes;
3. incluir a todos los participantes asignados aleatoriamente en el análisis.

Siempre es posible aplicar el primer principio. Sin embargo, a menudo el segundo es imposible debido a un abandono del estudio más allá del control de los autores del ensayo. Debido a ello, el tercer principio de realizar un análisis que incluya a todos los participantes sólo se puede seguir si se hacen suposiciones acerca de los valores faltantes (ver más adelante). Por lo tanto, muy pocos ensayos pueden realizar un verdadero análisis por IT sin hacer imputaciones, especialmente cuando el seguimiento es prolongado. En la práctica, los autores de los estudios pueden describir un análisis por IT incluso cuando hay algunos datos de desenlace faltantes. El término "IT" no tiene una definición clara y se utiliza de forma inconsistente en los estudios (Hollis 1999). Los revisores deberían utilizar el término solamente para implicar los tres principios anteriores, y deberían interpretar con precaución cualquier estudio que utilice el término sin aclaración.

Los revisores pueden encontrar que los análisis se describen como "intención de tratar modificada", lo que habitualmente significa que los participantes se excluyeron si no recibieron una cantidad mínima especificada de la intervención planificada. Este término también se utiliza en una variedad de formas por lo que los revisores siempre deberían buscar información acerca de quién se incluyó específicamente.

Es de señalar que pudiera ser posible realizar análisis que incluyan participantes excluidos por los autores del estudio (reinclusiones), si los motivos para las exclusiones se consideran inapropiados y los datos están disponibles para los revisores. Se les recomienda a los revisores que lo hagan, de ser posible y apropiado.

Las inquietudes acerca del sesgo debido a los datos de desenlace incompletos surgen principalmente a partir de consideraciones teóricas. Numerosos estudios empíricos han examinado si varios aspectos de los datos faltantes se asocian con la magnitud de las estimaciones del efecto. La mayoría no encontró evidencia clara de sesgo (Schulz 1995b, Kjaergard 2001, Balk 2002, Siersma 2007). Tierney y cols. observaron una tendencia a favor de la intervención experimental cuando los análisis se realizaron después que los autores excluyeron a los participantes, comparados con los análisis que incluyeron a todos los participantes (Tierney 2005). Sin embargo, no hay ejemplos significativos de análisis sesgados "por protocolo" (Melander 2003), y una revisión encontró estimaciones del efecto más exageradas con los análisis "por protocolo" que con los análisis por "IT" de los mismos ensayos (Porta 2007). La interpretación de los estudios empíricos es difícil debido a que las exclusiones se describen de forma deficiente, especialmente antes de 1996, en la era pre-CONSORT (Moher 2001a). Por ejemplo, Schulz observó que la *aparente* falta de exclusiones se asoció con tamaños del efecto más "beneficiosos", así como con menos probabilidades de una ocultación adecuada de la asignación (Schulz 1996). Por lo tanto, la imposibilidad de describir las exclusiones en los ensayos del estudio de Schulz puede haber sido un marcador de una realización deficiente del ensayo, en lugar de una verdadera falta de exclusiones.

Las investigaciones empíricas también han investigado si fue adecuada la forma en la cual los datos de desenlace incompletos se abordaron en los ensayos. Un estudio de las publicaciones de 71 ensayos de cuatro revistas médicas generales concluyeron que los datos faltantes son frecuentes y a menudo se tratan de forma inadecuada en el análisis estadístico (Wood 2004).

8.13.2 Evaluación del riesgo de sesgo a partir de los datos de desenlace incompletos

El riesgo de sesgo que surge de los datos de desenlace incompletos depende de numerosos factores, que incluyen la cantidad y la distribución entre los grupos de intervención, los motivos para que haya desenlaces faltantes, las diferencias probables en el desenlace entre los participantes con y sin datos, lo que hicieron los autores del estudio para tratar el problema en sus análisis, y el contexto clínico. Por lo tanto, no es posible formular una regla simple para evaluar si un estudio tiene un riesgo bajo o alto de sesgo. Las siguientes consideraciones pueden ayudar a los revisores a evaluar si los datos de desenlace

incompletos se pueden tratar de una forma que proteja contra el sesgo, cuando se utiliza la herramienta de la Colaboración (Sección 8.5).

A menudo se supone que una alta proporción de desenlaces faltantes, o una diferencia grande en las proporciones entre los grupos de intervención, es la causa principal de las inquietudes acerca del sesgo. Sin embargo, estas características por sí mismas no son suficientes para introducir sesgo. En esta sección se elaboran situaciones en las cuales se puede evaluar si un análisis tiene un riesgo bajo o alto de sesgo. Es fundamental considerar los motivos de los desenlaces faltantes, así como los números faltantes.

8.13.2.1 Bajo riesgo de sesgo debido a datos de desenlace incompletos

Para concluir que no hay datos de desenlace faltantes, los revisores deberían estar seguros de que los participantes incluidos en el análisis son exactamente los que se asignaron aleatoriamente en el ensayo. Si los números asignados aleatoriamente a cada grupo de intervención no se describen claramente, el riesgo de sesgo es poco claro. Como se señala anteriormente, no siempre es necesario considerar que los participantes asignados aleatoriamente, pero que posteriormente se consideró que no eran elegibles, tienen datos de desenlace faltantes.

Ejemplo (de bajo riesgo de sesgo): "Todos los participantes completaron el estudio y no hubo pérdidas durante el seguimiento, no hubo cambios en los grupos del ensayo y no hubo eventos adversos importantes".

Motivos aceptables para los datos faltantes

Es poco probable que la decisión de una persona sana de mudarse lejos de la localidad geográfica de un ensayo clínico esté conectada con su desenlace posterior. Para los estudios con una duración prolongada del seguimiento algunos abandonos por estos motivos son inevitables.

Para los estudios que describen datos de tiempo hasta el evento todos los participantes que no presentaron el evento de interés se consideran "censurados" en la fecha de su último seguimiento (no se conoce si el evento desenlace ocurrió después de que terminó el seguimiento). La consideración importante para este tipo de análisis es si se puede suponer que esta censura no está sesgada, es decir, que el efecto de la intervención (p.ej. evaluado mediante un índice de riesgo) en individuos censurados antes del final *planificado* del seguimiento es el mismo que el cociente de riesgos instantáneos en otros individuos. En otras palabras, no hay sesgo si la censura no se relaciona con el pronóstico.

Si hay datos de resultado faltantes en ambos grupos de intervención, pero los motivos para los mismos se detallan y están equilibrados entre los grupos, entonces no es de esperar que exista sesgo a menos de que los motivos tengan implicaciones diferentes en los grupos comparados. Por ejemplo, "negarse a participar" puede significar no estar dispuesto a hacer ejercicios en un grupo de ejercicio, mientras que negarse podría implicar insatisfacción con la recomendación de no hacer ejercicios en el otro grupo. En la práctica, la notificación incompleta de los motivos de los desenlaces faltantes puede impedir que los revisores hagan esta evaluación.

Posible repercusión de los datos faltantes sobre las estimaciones del efecto

La posible repercusión de los desenlaces dicotómicos faltantes depende de la frecuencia (o el riesgo) del desenlace. Por ejemplo, si el 10% de los participantes tiene desenlaces faltantes, entonces su posible repercusión sobre los resultados es mucho mayor si el riesgo del evento es del 10% que si es del 50%. La tabla siguiente ilustra la posible repercusión de los riesgos observados. A y B representan dos ensayos hipotéticos de 1000 participantes en

los cuales el 90% de los individuos son observados, y el cociente de riesgos entre estos 900 participantes observados es 1. Además, en ambos ensayos se supone que los participantes faltantes del grupo de intervención tienen un alto riesgo de presentar el evento (80%) y los del grupo control tienen un riesgo mucho menor (20%). La única diferencia entre los ensayos A y B es el riesgo entre los participantes observados. En el ensayo A el riesgo es del 50% y la repercusión de los datos faltantes, si se hubieran observado, es baja. En el ensayo B el riesgo es del 10% y la repercusión de los mismos datos faltantes, si se hubieran observado, es grande. Generalmente cuanto mayor es la proporción de participantes con datos faltantes con respecto a la de los participantes con eventos, mayor es la posibilidad de sesgo. En el ensayo A esta proporción fue 100/450 (0.2), mientras que en el ensayo B fue 100/90 (1.1).

	Número aleatorizado	Riesgo entre los observados	Datos observados	Riesgos extremos hipotéticos entre los participantes faltantes	Datos faltantes	Datos completos	Razón de riesgo basada en todos los participantes
Ensayo A							
Intervención	500	50%	225/450	80%	40/50	265/500	1,13
Control	500	50%	225/450	20%	10/50	235/500	
Ensayo B							
Intervención	500	10%	45/450	80%	40/50	85/500	1,55
Control	500	10%	45/450	20%	10/50	55/500	

La posible repercusión de los desenlaces continuos faltantes aumenta con la proporción de participantes con datos faltantes. También es necesario considerar el posible efecto de la intervención entre los participantes con resultados faltantes. La tabla siguiente ilustra la repercusión de las proporciones diferentes de desenlaces faltantes. A y B representan dos ensayos hipotéticos de 1000 participantes en los cuales la diferencia en la respuesta media entre la intervención y el control entre los participantes observados es 0. Además, en ambos ensayos se supone que los participantes faltantes en el brazo intervención tienen una media mayor y los del brazo control tienen una media menor. La única diferencia entre los ensayos A y B es el número de participantes faltantes. En el ensayo A el 90% de los participantes son observados y el 10% son faltantes, y la repercusión de los datos faltantes sobre la diferencia de medias observada es baja. En el ensayo B la mitad de los participantes son faltantes y la repercusión de los mismos datos faltantes sobre la diferencia de medias observada es grande.

	Número aleatorizado	Número observado	Media observada	Números faltantes	Media extrema hipotética entre los participantes faltantes	Media general (promedio ponderado)	Diferencia de medias basada en todos los participantes
Ensayo A							
Intervención	500	450	10	50	15	10.5	1
Control	500	450	10	50	5	9,5	
Ensayo B							
Intervención	500	250	10	250	15	12.5	5
Control	500	250	10	250	5	7,5	

8.13.2.2 Alto riesgo de sesgo debido a datos de desenlace incompletos

Motivos inaceptables para los datos faltantes

Una diferencia en la proporción de datos de desenlace incompletos entre los grupos es preocupante si la disponibilidad de los datos de desenlace está determinada por los desenlaces reales de los participantes. Por ejemplo, si es más probable que los participantes con los desenlaces clínicos más deficientes abandonen debido a los efectos adversos, y esto sucede principalmente en el grupo experimental, entonces las estimaciones del efecto estarán sesgadas a favor de la intervención experimental. La exclusión de participantes debido a "ineficacia" o "falta de mejoría" introducirá sesgo si los números excluidos no están equilibrados entre los grupos de intervención. Es de señalar que un resultado no significativo de una prueba estadística en presencia de pérdidas diferenciales no confirma la ausencia de sesgo, especialmente en los estudios pequeños.

Ejemplo (de alto riesgo de sesgo): "En un ensayo de sibutramina frente a placebo para tratar la obesidad, 13/35 se retiraron del grupo de sibutramina, 7 de ellos debido a falta de eficacia. Se retiraron 25/34 del grupo placebo, 17 debido a falta de eficacia. Un análisis "por intención de tratar" incluyó solamente a los restantes" (Cuellar 2000) (es decir, sólo 9 de 34 del grupo placebo) .

Incluso si los datos de desenlace incompletos están equilibrados en cuanto a los números entre los grupos, es posible introducir un sesgo si los motivos para los datos faltantes difieren. Por ejemplo, en un ensayo de una intervención experimental dirigida al abandono del hábito de fumar es posible que una proporción de los participantes en la intervención control abandonen el estudio debido a falta de entusiasmo por no recibir una intervención novedosa (y continúen con el hábito de fumar), y que una proporción similar del grupo de intervención experimental abandone el estudio debido a que no pudieron dejar de fumar.

Es posible que el enfoque común para tratar los desenlaces faltantes en los estudios de abandono del hábito de fumar (suponer que todos los que abandonaron el estudio continúan con el hábito) no siempre esté libre de sesgo. El ejemplo destaca la importancia de considerar *los motivos* de los datos de desenlace incompletos cuando se evalúa el riesgo de sesgo. En la práctica, a menudo no se conoce por qué la mayoría de los participantes abandonan el estudio, aunque un estudio empírico ha observado que 38 de 63 ensayos con datos faltantes proporcionaron información sobre los motivos (Wood 2004), y que esto es probable que mejore con el uso de la Declaración CONSORT (Moher 2001d).

Análisis "como se trató" (por protocolo)

Los participantes elegibles se deberían analizar en los grupos a los cuales se asignaron aleatoriamente, con independencia de la intervención que en realidad recibieron. Por lo tanto, en un estudio que compare cirugía con radioterapia para el tratamiento del cáncer de próstata localizado, los pacientes que rechacen la cirugía y seleccionen la radioterapia después de la asignación aleatoria se deberían incluir en el grupo de cirugía para el análisis. Esto se debe a que la propensión de los participantes a cambiar de grupo puede estar relacionada con el pronóstico, y en ese caso el intercambio entre los grupos de intervención introduce un sesgo de selección. Aunque este hecho tiene que ver específicamente con un análisis inapropiado y no con datos de desenlace incompletos, los estudios en los cuales se describió el análisis "como se trató" se deberían clasificar como alto riesgo de sesgo debido a datos de desenlace incompletos, a menos de que el número de intercambios sea demasiado pequeño para que represente diferencias importantes en la estimación del efecto de la intervención.

Un enfoque igualmente inapropiado para el análisis de un estudio es enfocarse solamente en los participantes que cumplieron con el protocolo. Un ejemplo ilustrativo lo proporciona un ensayo de un fármaco para disminuir los lípidos, el clofibrato (Coronary Drug Protect Research Gropu 1980). La mortalidad a los cinco años en 1103 hombres asignados a

clofibrato fue del 20,0% y en 2789 hombres asignados a placebo fue del 20,9% (p = 0,55). Los que cumplieron de forma adecuada con el protocolo en el grupo clofibrato tuvieron una menor mortalidad a los cinco años (15,0%) que los que no lo hicieron (24,6%). Sin embargo, se observó una diferencia similar entre "buenos cumplidores" y "cumplidores deficientes" en el grupo placebo (15,1% frente a 28,3%). Por lo tanto, el cumplimiento fue un marcador pronóstico en lugar de un modificador del efecto del clofibrato. Estos hallazgos muestran una grave dificultad al evaluar la eficacia de la intervención en los subgrupos por medio de las respuestas de los pacientes a las intervenciones. Debido a que no recibir la intervención puede ser más informativo que la no disponibilidad de los datos de desenlace, existe un alto riesgo de sesgo en los análisis limitados a los cumplidores, incluso con tasas bajas de datos incompletos.

8.13.2.3 Intentos de solucionar los datos faltantes en las publicaciones: imputación

Un enfoque frecuente pero potencialmente peligroso para tratar los datos de desenlace faltantes es imputar los resultados y tratarlos como si fueran mediciones reales (ver también Capítulo 16, Sección 16). Por ejemplo, los individuos con datos de desenlaces faltantes podrían asignarse al desenlace medio para su grupo de intervención, o al éxito o al fracaso del tratamiento. Dichos procedimientos pueden dar lugar a un sesgo grave y a intervalos de confianza demasiado estrechos. Una variante de esto, cuya validez es más difícil de evaluar, es el uso de "la última observación realizada" (UOR). Aquí se asume la medida de desenlace más recientemente observada para todos los momentos posteriores a las evaluaciones de desenlace (Lachin 2000, Unnebrink 2001). Los procedimientos de la UOR también pueden dar lugar a sesgos graves. Por ejemplo, en un ensayo de un fármaco para una afección degenerativa como la enfermedad de Alzheimer el abandono puede estar relacionado con los efectos secundarios del fármaco. Como los resultados tienden a deteriorarse con el tiempo, el uso de la UOR sesgará la estimación del efecto a favor del fármaco. Por el contrario, el uso de la UOR podría ser apropiado si la mayoría de las personas para las cuales se utiliza este método en el análisis de los desenlaces tienen una medición real relativamente reciente.

Existen muchas publicaciones sobre los métodos estadísticos para tratar los datos faltantes de una forma válida (ver Capítulo 16, Sección 16.1). Hay relativamente pocas aplicaciones prácticas de estos métodos en los ensayos clínicos (Wood 2004). Se recomienda asesoría estadística si los revisores tienen que utilizarlos. Un buen punto de partida para aprender sobre ésto es www.missingdata.org.uk.

8.14 Notificación selectiva de los desenlaces

8.14.1 Justificación para las inquietudes acerca del sesgo

La notificación selectiva de los desenlaces se ha definido como la selección de un subgrupo de las variables originales registradas, basadas en los resultados, para su inclusión en la publicación de los ensayos (Hutton 2000); ver también Capítulo 10 (Sección 10.2.2.5). La principal preocupación es que los resultados estadísticamente no significativos podrían no ser revelados en la publicación de forma selectiva. Hasta hace poco la evidencia publicada de la notificación selectiva de los desenlaces fueron limitadas. Inicialmente fueron unos pocos estudios de casos. Luego un pequeño estudio de una cohorte completa de solicitudes aprobadas por un único Comité de Ética de Investigación Local encontró que el desenlace primario se señaló en sólo seis de los protocolos de las 15 publicaciones obtenidas. Ocho protocolos hicieron alguna referencia a un análisis planificado, pero siete de las publicaciones no siguieron este plan de análisis (Hahn 2002). La notificación de los desenlaces dentro del estudio fue evidente o se sospechó en varios ensayos incluidos en una revisión de una cohorte de cinco metanálisis en la *Base de Datos Cochrane de Revisiones Sistemáticas* (*Cochrane Database of Systematic Reviews)* (Williamson 2005a).

La evidencia empírica convincente de la existencia de sesgo de notificación selectiva dentro de los estudios provienen de tres estudios recientes. En el primer estudio (Chan 2004a) se identificaron 102 ensayos con 122 publicaciones y 3736 desenlaces. En general se notificaron de forma incompleta (una mediana de) 38% de desenlaces de eficacia y 50% de desenlaces de seguridad por ensayo de grupo paralelo, es decir, con información insuficiente para incluirlos en un metanálisis. Los desenlaces estadísticamente significativos tuvieron probabilidades mayores de estar descritos completamente cuando se compararon con los resultados no significativos, para los datos de eficacia (*odds ratio* agrupado 2,4; IC del 95%: 1,4 a 4,0) y de daños (4,7; 1,8 a 12). Además, cuando se compararon las publicaciones con los protocolos, el 62% de los ensayos tuvo al menos un desenlace primario cambiado, introducido u omitido. Un segundo estudio de 48 ensayos financiado por los *Canadian Institutes of Health Research* encontró resultados muy similares (Chan 2004b). Un tercer estudio que incluyó una revisión retrospectiva de 519 publicaciones de ensayos y una encuesta de seguimiento a los autores comparó los resultados presentados con los desenlaces mencionados en la sección "Métodos" del mismo artículo (Chan 2005). Como promedio, alrededor del 20% de los desenlaces medidos en los ensayos de grupos paralelos fueron descritos de forma incompleta. Dentro de los ensayos estos desenlaces tuvieron probabilidades mayores de ser estadísticamente no significativos comparados con los desenlaces notificados de forma completa (*odds ratio* 2,0 [1,6 a 2,7] para los resultados de eficacia; 1,9 [1,1 a 3,5] para los resultados de daños). Estos tres estudios indican un *odds ratio* de cerca de 2,4 asociado con la notificación selectiva de los desenlaces, el cual corresponde, por ejemplo, a cerca del 50% de los desenlaces no significativos publicados, comparado con el 72% de los desenlaces significativos.

En los tres estudios se les preguntó a los autores si había desenlaces no publicados, si éstos mostraron diferencias significativas y por qué estos desenlaces no se habían publicado. Los motivos más frecuentes para no publicar los resultados fueron "falta de importancia clínica" o falta de significación estadística. Por lo tanto, es probable que los metanálisis que excluyen desenlaces no publicados sobrestimen los efectos de la intervención. Además, con frecuencia los autores omitieron mencionar la existencia de desenlaces no publicados incluso cuando éstos se habían mencionado en el protocolo o la publicación.

Estudios recientes han encontrado resultados similares (Ghersi 2006, von Elm 2006). En un tipo de estudio diferente, el efecto en los metanálisis fue mayor cuando un número menor de los ensayos disponibles contribuyeron con datos para el metanálisis (Furukawa 2007). Este hallazgo también indica que los resultados pueden haber sido selectivamente no revelados por los autores de los ensayos sobre la base de la magnitud del efecto.

Es probable que exista un sesgo asociado con la notificación selectiva de diferentes medidas de una misma característica. En los ensayos de tratamiento de la esquizofrenia se ha observado que un efecto de la intervención es más probable cuando se utilizan escalas de clasificación no publicadas, en lugar de publicadas (Marshall 2000). Los autores plantearon la hipótesis de que es menos probable que se publiquen los datos de escalas no publicadas cuando no hay significación estadística o que, después del análisis, se hayan retirado ítems desfavorables para crear un aparente efecto beneficioso.

En muchas revisiones sistemáticas sólo es posible incluir unos pocos estudios elegibles en un metanálisis para un desenlace específico debido a que no se detalló la información necesaria en los otros estudios. Aunque ese desenlace puede no haberse evaluado en algunos estudios, casi siempre existe un riesgo de notificación sesgado en varios de ellos. Los revisores deberían considerar si un desenlace se obtuvo pero no se describió o simplemente no se obtuvo.

La notificación selectiva de los desenlaces puede surgir de diferentes maneras; algunas afectan el estudio en general (punto 1 a continuación) y otras se relacionan con desenlaces específicos (puntos 2 al 6 a continuación):

1. Omisión selectiva de los desenlaces de las publicaciones: Es posible que solamente algunos de los desenlaces analizados se incluyan en la publicación. Si esta selección se hace según los resultados, en particular la significación estadística, es probable que las estimaciones metanalíticas correspondientes estén sesgadas.

2. Selección selectiva de los datos para un desenlace: Para un desenlace específico puede haber diferentes puntos temporales en los cuales se midió, o se pueden haber utilizado diferentes instrumentos para medir el desenlace en el mismo punto temporal (p.ej. escalas diferentes o evaluadores diferentes). Por ejemplo, en un ensayo de osteoporosis hubo 12 grupos diferentes de datos a seleccionar para la estimación del contenido mineral óseo. La diferencia de medias estandarizada para estas 12 posibilidades varió entre −0,02 y 1,42 (Gøtzsche 2007). Si los autores del estudio realizan la selección según estos resultados, entonces la estimación metanalítica estará sesgada.

3. Notificación selectiva de los análisis utilizando los mismos datos: A menudo hay varias formas diferentes con las cuales es posible analizar un resultado. Por ejemplo, los desenlaces continuos como la reducción de la presión sanguínea se pueden analizar como variables continuas o dicotómicas, y además existe la posibilidad de seleccionar múltiples puntos de corte. Otra alternativa frecuente para el análisis son las puntuaciones finales frente a los cambios a partir de los valores iniciales (Williamson 2005b). Cambiar una comparación planificada de los valores finales por una comparación de los cambios a partir del valor inicial debido a un desequilibrio en los valores iniciales observados en realidad introduce sesgo en lugar de eliminarlo (como podrían suponer los autores del estudio) (Senn 1991, Vickers 2001).

4. Notificación selectiva de subgrupos de datos: La notificación selectiva puede ocurrir si los datos de desenlace se pueden subdividir, por ejemplo, al seleccionar subescalas de una escala de medición completa o un subgrupo de eventos. Por ejemplo, es posible identificar infecciones micóticas al inicio o un par de días después de la asignación aleatoria, o pueden ser las llamadas infecciones micóticas “emergentes” que se detectan algunos días después de la asignación aleatoria, y la selección de un subgrupo de estas infecciones puede dar lugar a sesgo de notificación (Jørgensen 2006, Jørgensen 2007).

5. Subnotificación selectiva de los datos: Es posible que algunos resultados se describan pero los detalles de los datos no sean suficientes para incluirlos en un metanálisis. En algunas ocasiones se relaciona de forma explícita con el resultado, por ejemplo, cuando se detalla solamente como “no significativo” o “p > 0,05”.

Hay otras formas de notificación selectiva que no se mencionan aquí; éstas incluyen la notificación selectiva de los análisis de subgrupos o los análisis ajustados, y la presentación de los resultados del primer período en los ensayos cruzados (Williamson 2005a). Además, en ocasiones las descripciones de los desenlaces como “primarios”, “secundarios”, etc. se modifican de forma retrospectiva a la luz de los hallazgos (Chan 2004a, Chan 2004b). Generalmente este aspecto por sí mismo no debería ser motivo de inquietud para los revisores (que no toman en cuenta cómo se clasifican los desenlaces en cada estudio), siempre que no influya en qué resultados se publican.

8.14.2 Evaluación del riesgo de sesgo de la notificación selectiva de los desenlaces

Aunque la posibilidad de sesgo de publicación *entre estudios* se puede examinar solamente al considerar un grupo completo de estudios (ver Capítulo 10), la posibilidad de la notificación selectiva de los desenlaces *dentro del estudio* se puede examinar para cada estudio incluido en una revisión sistemática. Las siguientes consideraciones pueden ayudar a los revisores a evaluar si la descripción del desenlace es suficientemente completo y transparente para proteger contra el sesgo al utilizar la herramienta de la Colaboración (ver Capítulo 8.5).

Hasta el momento los métodos estadísticos para detectar la notificación selectiva dentro del estudio no están bien desarrollados. Sin embargo, hay otras formas de detectar dicho sesgo aunque es probable que una evaluación cuidadosa necesite un trabajo intenso. Si el protocolo está disponible es posible comparar los desenlaces del protocolo y de la publicación. De no ser así, entonces los desenlaces enumerados en la sección "Métodos" de un artículo se pueden comparar con los de la sección en la que se describen los resultados. Si se mencionan los resultados no significativos pero no se detallan de forma adecuada, es probable que exista un sesgo en el metanálisis. También se puede solicitar información adicional a los autores del estudio, aunque se debería tener en cuenta que dicha información podría ser poco fiable (Chan 2004a).

Algunas diferencias entre el protocolo y la publicación se pueden explicar por cambios admisibles que se han hecho al protocolo. Aunque dichos cambios se deberían detallar en las publicaciones, ninguno de los 150 estudios en las dos muestras de Chan y cols. lo hizo (Chan 2004a, Chanb).

Los revisores deberían buscar de forma intensiva la evidencia de que los autores de los estudios obtuvieron un escaso número de los desenlaces clave que habitualmente se miden en el área de interés, y comunicar qué estudios describen datos sobre ellos y cuáles no. Los revisores deberían considerar los *motivos* de por qué podría haber datos faltantes para un metanálisis (Williamson 2005b). Los métodos para buscar esta evidencia no están bien establecidos, pero aquí se describen algunas estrategias posibles.

Un primer paso útil es construir una matriz que indique qué resultados se registraron en qué estudio, por ejemplo, con filas para los estudios y columnas para los desenlaces. También se puede indicar la notificación completa y la incompleta. Esta matriz mostrará a los revisores qué estudios no describieron los desenlaces presentados por la mayoría de los otros estudios.

PubMed, otras importantes bases de datos de referencias e Internet se deberían explorar en busca de un protocolo del estudio; en pocas ocasiones se proporciona la dirección en la web de la publicación de un estudio. Alternativamente, y más a menudo en el futuro ya que será más frecuente el registro obligatorio de los ensayos, puede estar disponible en un registro de ensayos una descripción detallada del mismo. Los resúmenes de las presentaciones relacionadas con el estudio pueden contener información acerca de desenlaces que no se mencionan posteriormente en las publicaciones. Además, los revisores deberían examinar cuidadosamente la sección "Métodos" de los artículos publicados en busca de detalles de los resultados que se evaluaron.

Es de particular interés la información faltante que al parecer es seguro que se registró. Por ejemplo, es de esperar que algunas mediciones aparezcan juntas, como la presión sistólica y diastólica, por lo que uno se preguntaría por qué se describe una sola. Un ejemplo alternativo es un estudio que presenta la proporción de participantes cuyo cambio en una variable continua excedió cierto umbral; los investigadores deberían haber tenido acceso a los datos brutos y por lo tanto pueden haber mostrado los resultados como medias y desviaciones estándar de los cambios. Williamson y cols. dan numerosos ejemplos, incluida una revisión Cochrane en la cual nueve ensayos documentaron el desenlace "fracaso del tratamiento" pero sólo cinco notificaron sobre mortalidad. Incluso la mortalidad fue parte de la definición de fracaso del tratamiento, por lo que los datos deberían haberse obtenido en los cuatro ensayos faltantes del análisis de mortalidad. El sesgo se sospechó por la marcada diferencia en los resultados para fracaso del tratamiento para los ensayos con o sin notificación por separado de la mortalidad (Williamson 2005a).

Cuando existe sospecha o evidencia directa de la notificación selectiva de los desenlaces, es conveniente solicitar a los autores del estudio información adicional. Por ejemplo, se les puede solicitar a los autores que proporcionen el protocolo del estudio y la información completa sobre los desenlaces que no se documentaron de forma adecuada. Además, para los

desenlaces mencionados en el protocolo pero sobre los cuales no se proporcionó información, se les puede pedir que aclaren si dichas medidas de desenlace realmente se analizaron, y de ser así, que proporcionen los datos.

En general no se recomienda tratar de "ajustar por" el sesgo de notificación en el metanálisis principal. El análisis de sensibilidad es un mejor enfoque para investigar la posible repercusión de la notificación selectiva de los desenlaces (Hutton 2000, Williamson 2005a).

La evaluación del riesgo de sesgo debido a la notificación selectiva de los desenlaces se debería realizar para el estudio en general, en lugar de para cada desenlace. Aunque puede quedar claro para un estudio determinado que algunos desenlaces específicos pueden estar sujetos a la notificación selectiva mientras que otros no, se recomienda el enfoque a nivel de estudio porque no es práctico enumerar todos los desenlaces por completo en la tabla "Riesgo de sesgo". Se debería utilizar la sección "Descripción" de esta herramienta (ver Sección 8.5.2) para describir los desenlaces para los cuales existe una determinada evidencia de sesgo de notificación selectiva (o incompleta). La evaluación a nivel de estudio proporciona una valoración de la susceptibilidad general del estudio al sesgo de notificación selectiva.

8.15 Otras amenazas potenciales a la validez

8.15.1 Justificación para las inquietudes acerca del sesgo

Los dominios anteriores (generación de la secuencia, ocultación de la asignación, cegamiento, datos de resultado incompletos y notificación selectiva de los resultados) se relacionan con importantes fuentes potenciales de sesgo en los ensayos clínicos en todas las áreas de la atención sanitaria. Sin embargo, más allá de estos dominios específicos los revisores deberían estar al tanto de otros problemas que pueden surgir con respecto a la posibilidad de sesgo. El séptimo dominio en la herramienta de evaluación "Riesgo de sesgo" es un "todo incluido" de otras fuentes de sesgo. Para las revisiones en algunas áreas de interés existen muchos otros interrogantes adicionales que se deberían plantear en todos los estudios. En particular, algunos diseños de estudios merecen consideración especial cuando se encuentran. Si se prevé la existencia de diseños de estudios específicos (p.ej. ensayos cruzados [*crossover*] o tipos de estudios no aleatorizados) pueden plantearse interrogantes adicionales relacionadas con el riesgo de sesgo en estos tipos de estudios. La evaluación del riesgo de sesgo en los estudios no aleatorizados se aborda en el Capítulo 13, y el riesgo de sesgo para los ensayos aleatorizados grupales, los ensayos cruzados y los ensayos con múltiples grupos de intervención se abordan en el Capítulo 16. Además, es posible identificar algunos problemas importantes imprevistos con estudios específicos durante el desarrollo de la revisión sistemática o el metanálisis. Por ejemplo, un ensayo puede tener un desequilibrio sustancial en las características de los participantes al inicio. En las secciones siguientes se analizan varios ejemplos.

8.15.1.1 Riesgo de sesgo en diseños específicos

La principal inquietud acerca del riesgo de sesgo en los estudios no aleatorizados es el sesgo de selección en forma de diferencias en el tipo de participantes entre los grupos de intervención experimental y control. Los revisores deberían consultar la discusión completa en el Capítulo 13 (Sección 13.5). Las inquietudes principales acerca del riesgo de sesgo en los ensayos aleatorizados grupales son: (i) sesgo de reclutamiento (reclutamiento diferencial de los participantes en los grupos de diferentes intervenciones); (ii) desequilibrio inicial; (iii) pérdida de los grupos; (iv) análisis incorrecto; y (v) comparabilidad con los ensayos aleatorizados individuales. Las principales inquietudes acerca del riesgo de sesgo en los ensayos cruzados son: (i) si el diseño cruzado es adecuado; (ii) si existe un efecto de arrastre (*carry-over*); (iii) si sólo están disponibles los datos del primer período; (iv) análisis

incorrecto; y (v) comparabilidad de los resultados con los de los ensayos de grupos paralelos. Estos problemas se discuten en detalle en el Capítulo 16 (Secciones 16.3 y 16.4). El riesgo de sesgo en los estudios con más de dos grupos de intervención también se discute en el Capítulo 16 (Sección 16.5).

8.15.1.2 Desequilibrio inicial

El desequilibrio inicial en los factores que se relacionan firmemente con las medidas de resultado puede provocar sesgo en la estimación del efecto de la intervención. Esto puede suceder sólo por azar, pero el desequilibrio también puede surgir a partir de la asignación no aleatoria (no oculta) de las intervenciones. Algunas veces los autores pueden excluir algunos individuos asignados al azar, lo que provoca desequilibrio en las características de los participantes en los diferentes grupos de intervención. La generación de la secuencia, la falta de ocultación de la asignación o la exclusión de participantes se deberían abordar mediante ítems específicos para estos problemas en la herramienta. Se debería señalar si se observa otro desequilibrio inicial inexplicable, suficiente para dar lugar a una exageración importante en los efectos. Las pruebas de desequilibrio inicial no tienen valor en los ensayos verdaderamente aleatorizados, pero valores de p muy pequeños pudieran indicar sesgo en la asignación de la intervención.

Ejemplo (de alto riesgo de sesgo): Un ensayo de captopril versus un antihipertensivo convencional tuvo desequilibrios pequeños pero altamente significativos en la talla, el peso, la PA sistólica y diastólica: $p = 10^{-4}$ a 10^{-18} (Hansson 1999). Dicho desequilibrio indica fracaso de la asignación al azar (que se realizó mediante sobres cerrados) en algunos centros (Peto 1999).

8.15.1.3 Asignación en bloques en ensayos no cegados

Algunas combinaciones de métodos para la generación de la secuencia, la ocultación de la asignación y el cegamiento actúan juntas para crear un riesgo de selección en la asignación a las intervenciones. Una combinación específica es el uso de la asignación aleatoria en bloques en un ensayo no cegado, o en un ensayo cegado en el que se rompe el cegamiento, por ejemplo debido a los efectos secundarios característicos. Cuando se utiliza la asignación en bloques, y cuando las asignaciones se revelan posteriormente a la persona responsable del reclutamiento en el ensayo, algunas veces es posible predecir las asignaciones siguientes. Esto es especialmente el caso cuando los bloques tienen un tamaño fijo y no se dividen entre los múltiples centros de reclutamiento. Esta capacidad de predecir las asignaciones siguientes puede ocurrir incluso si la ocultación de la asignación es adecuada según los criterios indicados en la Tabla 8.5.d (Berger 2005).

8.15.1.4 Actividad diagnóstica diferencial

Las evaluaciones de resultado pueden estar sesgadas a pesar del cegamiento efectivo. En particular, el aumento de la actividad diagnóstica puede dar lugar a un aumento del diagnóstico de casos de enfermedad verdaderos pero asintomáticos. Por ejemplo, muchas ulceras gástricas no dan síntomas y no tienen relevancia clínica, pero estos casos se pueden detectar con más frecuencia por gastroscopia en pacientes que reciben un fármaco que provoca molestias gástricas inespecíficas que por lo tanto provocan más gastroscopias. De manera similar, si un fármaco provoca diarrea, pudiera causar más exámenes rectales digitales y, por lo tanto, también un aumento de la detección de casos de cáncer prostático asintomáticos. Obviamente, la evaluación de los efectos beneficiosos también puede estar sesgada mediante el mismo mecanismo. Asimismo, las intervenciones pueden dar lugar a una actividad diagnóstica diferencial, por ejemplo, si la intervención experimental es la visita de una enfermera al domicilio del paciente, y la intervención control es ninguna visita.

8.15.1.5 Otros ejemplos de sesgos potenciales

La siguiente lista de otras posibles fuentes de sesgo en un estudio clínico puede ayudar a detectar problemas adicionales:

- La realización del estudio está afectada por los resultados intermedios (p.ej. el reclutamiento adicional de participantes de un subgrupo que muestra más beneficios).
- Existe una desviación del protocolo del estudio en una manera que no refleja la práctica clínica (p.ej. aumento gradual *post hoc* de la dosis hasta niveles exagerados).
- Existe la administración previa a la asignación al azar de una intervención que pudiera mejorar o disminuir el efecto de una intervención aleatoria posterior.
- Administración inapropiada de una intervención (o cointervención).
- Contaminación (p.ej. participantes combinando fármacos).
- Aparición de "sesgo nulo" debido a intervenciones que no se administran suficientemente bien o criterios de inclusión demasiado amplios para los participantes (Woods 1995).
- Se utiliza un instrumento que no tiene la sensibilidad para medir los resultados (lo cual puede dar lugar a la sobrestimación de los efectos beneficiosos y perjudiciales).
- Notificación selectiva de subgrupos.
- Fraude.

La influencia inapropiada de los financiadores (p.ej. en un estudio empírico, más de la mitad de los protocolos de los ensayos iniciados por la industria señalaron que el patrocinador era el propietario de los datos o debía aprobar el texto, o ambos; ninguna de estas limitaciones se señaló en las publicaciones de los ensayos (Gøtzsche 2006).Es importante que la información sobre los intereses particulares se recoja y presente cuando sea relevante. Sin embargo, los revisores deberían proporcionar esta información en la tabla de "Características de los estudios incluidos" (ver Sección 11.2.2). La tabla "Riesgo de sesgo" debería utilizarse para evaluar aspectos específicos de metodología que pudieran haber sido influenciados por los intereses particulares y que podrían conducir directamente al riesgo de sesgo. Cabe destacar que algunas decisiones que pueden estar influenciadas por aquellos con un interés particular, como el escoger una dosis especialmente baja de un fármaco comparador, deberían abordarse como una fuente de heterogeneidad en lugar de utilizar la herramienta "Riesgo de sesgo", puesto que no impactan directamente en la validez interna de los resultados.

8.15.2 Evaluación del riesgo de sesgo de otras fuentes

Más adelante se proporcionan algunas pautas para determinar los temas adecuados para las evaluaciones de "otras fuentes de sesgo". En particular, estos temas adecuados deberían constituir fuentes potenciales de sesgo y no fuentes de imprecisión, fuentes de diversidad (heterogeneidad) o medidas de calidad de la investigación que no se relacionan con el sesgo. Los temas tratados en este dominio de la herramienta incluyen principalmente los ejemplos brindados en la Sección 8.15.1. Sin embargo, más allá de estos aspectos específicos los revisores deberían tener en cuenta aspectos específicos de los estudios que pueden hacer surgir inquietudes acerca de la posibilidad de sesgo, y deberían formular consideraciones acerca de ellos en este dominio de la herramienta. Las siguientes consideraciones pueden ayudar a los revisores a evaluar si un estudio está libre de otras fuentes de sesgo cuando utilicen la herramienta de la Colaboración (Sección 8.5).

De ser posible, un protocolo de revisión debería preespecificar cualquier interrogante a abordar, lo cual daría lugar a ítems separados en la tabla "Riesgo de sesgo". Por ejemplo, si los ensayos cruzados son el diseño de estudio habitual para la interrogante que se va a abordar en la revisión, entonces es necesario formular por adelantado preguntas específicas relacionadas con el sesgo en los estudios cruzados.

Los aspectos incluidos en la herramienta para el riesgo de sesgo deberían ser una fuente potencial de sesgo, y no sólo una causa de *imprecisión* (ver Sección 8.2), lo cual se aplica a aspectos evaluados en el dominio "Otras fuentes de sesgo". Una fuente potencial de sesgo

debería ser capaz de cambiar la magnitud de la estimación del efecto, mientras que las fuentes de imprecisión afectan solamente a la incertidumbre de la estimación (es decir, su intervalo de confianza). Los posibles factores que afectan la precisión de una estimación incluyen la variabilidad tecnológica (p.ej. el error de medición) y la variabilidad del observador.

Como la herramienta aborda solamente los sesgos internos, cualquier aspecto cubierto por este dominio debería ser una fuente potencial de sesgo interno y no una fuente de *diversidad*. Las posibles causas de diversidad incluyen las diferencias en la dosis del fármaco, la duración del seguimiento y las características de los participantes (p.ej. edad, estadio de la enfermedad). Los estudios pueden seleccionar dosis que favorecen el fármaco experimental con respecto al fármaco control. Por ejemplo, a menudo los fármacos más antiguos se sobredosifican (Safer 2002) o se pueden administrar en circunstancias claramente subóptimas que no reflejan la práctica clínica (Johansen 2000, Jørgensen 2007). Alternativamente, es posible que los pacientes se recluten de forma selectiva para su inclusión en un estudio sobre la base de una "respuesta" previamente demostrada a la intervención experimental. Es importante que estas selecciones sesgadas se aborden en las revisiones Cochrane. Aunque es posible que no se hayan incluido en la herramienta "Riesgo de sesgo" descrita en el presente Capítulo, se pueden haber abordado en algunas ocasiones en el análisis (p.ej. por análisis de subgrupos y metarregresión) y se deberían considerar en la clasificación e interpretación de la evidencia en una tabla de "Resumen de los hallazgos" (ver Capítulos 11 y 12).

Es posible realizar muchas evaluaciones acerca del diseño y la realización de un ensayo clínico, pero no todas se asocian con el sesgo. Las medidas de "calidad" solas a menudo se asocian firmemente con aspectos que pudieran introducir sesgo. Sin embargo, los revisores se deberían centrar en los mecanismos que dan lugar a sesgo en lugar de los descriptores de los estudios que reflejan solamente la "calidad". Algunos ejemplos de indicadores de "calidad" que no se deberían evaluar dentro de este dominio incluyen los criterios relacionados con la aplicabilidad, "generalizabilidad" o la "validez externa" (incluidos los señalados anteriormente), los criterios relacionados con la precisión (p.ej. tamaño de muestra o uso de un cálculo del tamaño de muestra [poder estadístico]), estándares para su publicación, y criterios éticos (p.ej. si el estudio tuvo la aprobación ética o los participantes dieron su consentimiento informado). Estos factores pueden ser importantes y se presentarán en la tabla "Características de los estudios incluidos" o en "Tablas adicionales" (ver Capítulo 11).

Finalmente, para evitar el doble recuento, las fuentes potenciales de sesgo no se deberían incluir como "sesgo de otras fuentes" si se incluyen de forma más apropiada en los dominios anteriores de la herramienta. Por ejemplo, en la enfermedad de Alzheimer los pacientes se deterioran significativamente con el tiempo durante el ensayo. Generalmente los efectos del tratamiento son pequeños y los tratamientos tienen una marcada toxicidad. Es muy difícil tratar de forma satisfactoria las pérdidas de participantes. Es probable que los que reciben el tratamiento abandonen más temprano debido a efectos adversos o muerte, y por lo tanto las mediciones en estas personas tienden a ser más tempranas en el estudio y favorecerán la intervención. A menudo es difícil monitorizar de forma continua estos participantes para realizar un análisis de todos los participantes asignados al azar. Este aspecto, aunque pudiera parecer inicialmente una causa de sesgo específica del tema, sería más apropiado incluirlo en "Datos de resultado incompletos".

8.16 Información del capítulo

Editores: Julian PT Higgins, Douglas G Altman y Jonathan AC Sterne del Grupo Cochrane de Métodos Estadísticos y del Grupo Cochrane de Métodos de Sesgo.

La versión en inglés de este capítulo debería citarse como: Higgins JPT, Altman DG, Sterne JAC (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0

(updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Autores que contribuyeron: Doug Altman, Gerd Antes, Peter Gøtzsche, Julian Higgins, Peter Jüni, Steff Lewis, David Moher, Andy Oxman, Ken Schulz, Jonathan Sterne y Simon Thompson.

Agradecimientos: El material de este capítulo fue desarrollado por un grupo de trabajo constituido por Doug Altman (codirector), Gerd Antes, Chris Cates, Mike Clarke, Jon Deeks, Peter Gøtzsche, Julian Higgins (codirector), Sally Hopewell, Peter Jüni (grupo principal), Steff Lewis, Philippa Middleton, David Moher (grupo principal), Andy Oxman, Ken Schulz (grupo principal), Nandi Siegfried, Jonathan Sterne y Simon Thompson. La revisión de la herramienta realizada en 2010 surgió de un proyecto de evaluación que involucró a Doug Altman (grupo principal), Elaine Beller, Sally Bell-Syer, Isabelle Boutron, Chris Cates, Rachel Churchill, June Cody, Jonathan Cook, Christian Gluud, Julian Higgins (grupo principal), Sally Hopewell, Hayley Jones, Peter Jűni, Monica Kjeldstrøm, Toby Lasserson, Allyson Lipp, Lara Maxwell, Joanne McKenzie, David Moher (grupo principal), Craig Ramsey, Barney Reeves, Jelena Savović (codirector), Jonathan Sterne (codirector), David Tovey, Lucy Turner and Laura Weeks (grupo principal). Se agradece a Hilda Bastian, Rachelle Buchbinder, Iain Chalmers, Miranda Cumpston, Sally Green, Peter Herbison, Victor Montori, Hannah Rothstein, Georgia Salanti, Guido Schwarzer, Ian Shrier, Jayne Tierney, Ian White y Paula Williamson por sus útiles comentarios. Para detalles del Grupo Cochrane de Métodos Estadísticos, ver Capítulo 9 (Recuadro 9.8.a), y para el Grupo Cochrane de Métodos de Sesgo, ver Capítulo 10 (Recuadro 10.5.a).

8.17 Referencias

Als-Nielsen 2004
Als-Nielsen B, Gluud LL, Gluud C. Methodological quality and treatment effects in randomized trials: a review of six empirical studies. *12th Cochrane Colloquium*, Ottawa (Canada), 2004.

Altman 1999
Altman DG, Bland JM. How to randomize. *BMJ* 1999; 319: 703-704.

Balk 2002
Balk EM, Bonis PAL, Moskowitz H, Schmid CH, Ioannidis JPA, Wang C, Lau J. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002; 287: 2973-2982.

Bellomo 2000
Bellomo R, Chapman M, Finfer S, Hickling K, Myburgh J. Low-dose dopamine in patients with early renal dysfunction: a placebo-controlled randomised trial. Australian and New Zealand Intensive Care Society (ANZICS) Clinical Trials Group. *The Lancet* 2000; 356: 2139-2143.

Berger 2003
Berger VW, Ivanova A, Knoll MD. Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. *Statistics in Medicine* 2003; 22: 3017-3028.

Berger 2005
Berger VW. Quantifying the magnitude of baseline covariate imbalances resulting from selection bias in randomized clinical trials. *Biometrical Journal* 2005; 47: 119-127.

Berlin 1997
Berlin JA. Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. *The Lancet* 1997; 350: 185-186.

Boutron 2005
Boutron I, Estellat C, Ravaud P. A review of blinding in randomized controlled trials found results inconsistent and questionable. *Journal of Clinical Epidemiology* 2005; 58: 1220-1226.

Boutron 2006
Boutron I, Estellat C, Guittet L, Dechartres A, Sackett DL, Hróbjartsson A, Ravaud P. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLOS Medicine* 2006; 3: 1931-1939.

Brightling 2000
Brightling CE, Monteiro W, Ward R, Parker D, Morgan MD, Wardlaw AJ, Pavord ID. Sputum eosinophilia and short-term response to prednisolone in chronic obstructive pulmonary disease: a randomised controlled trial. *The Lancet* 2000; 356: 1480-1485.

Brown 2005
Brown S, Thorpe H, Hawkins K, Brown J. Minimization: reducing predictability for multi-centre trials whilst retaining balance within centre. *Statistics in Medicine* 2005; 24: 3715-3727.

Chan 2004a
Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004; 291: 2457-2465.

Chan 2004b
Chan AW, Krleža-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *Canadian Medical Association Journal* 2004; 171: 735-740.

Chan 2005
Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005; 330: 753.

Coronary Drug Project Research Group 1980
Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *New England Journal of Medicine* 1980; 303: 1038-1041.

Cuellar 2000
Cuellar GEM, Ruiz AM, Monsalve MCR, Berber A. Six-month treatment of obesity with sibutramine 15 mg; a double-blind, placebo-controlled monocenter clinical trial in a Hispanic population. *Obesity Research* 2000; 8: 71-82.

de Gaetano 2001
de Gaetano G. Low-dose aspirin and vitamin E in people at cardiovascular risk: a randomised trial in general practice. Collaborative Group of the Primary Prevention Project. *The Lancet* 2001; 357: 89-95.

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. *Journal of Clinical Epidemiology* 1992; 45: 255-265.

Devereaux 2001
Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, Bhandari M, Guyatt GH. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001; 285: 2000-2003.

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 Clinical Trials* 1990; 11: 339-352.

Fergusson 2002
Fergusson D, Aaron SD, Guyatt G, Hébert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ* 2002; 325: 652-654.

Fergusson 2004
Fergusson D, Glass KC, Waring D, Shapiro S. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. *BMJ* 2004; 328: 432.

Furukawa 2007
Furukawa TA, Watanabe N, Omori IM, Montori VM, Guyatt GH. Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. *JAMA* 2007; 297: 468-470.

Ghersi 2006
Ghersi D, Clarke M, Simes J. Selective reporting of the primary outcomes of clinical trials: a follow-up study. *14th Cochrane Colloquium*, Dublin (Ireland), 2006.

Gøtzsche 1996
Gøtzsche PC. Blinding during data analysis and writing of manuscripts. *Controlled Clinical Trials* 1996; 17: 285-290.

Gøtzsche 2006
Gøtzsche PC, Hróbjartsson A, Johansen HK, Haahr MT, Altman DG, Chan AW. Constraints on publication rights in industry-initiated clinical trials. *JAMA* 2006; 295: 1645-1646.

Gøtzsche 2007
Gøtzsche PC, Hróbjartsson A, Maric K, Tendal B. Data extraction errors in meta-analyses that use standardized mean differences. *JAMA* 2007; 298: 430-437.

Greenland 2001
Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics* 2001; 2: 463-471.

Guyatt 2008
Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336: 924-926.

Haahr 2006
Haahr MT, Hróbjartsson A. Who is blinded in randomised clinical trials? A study of 200 trials and a survey of authors. *Clinical Trials* 2006; 3: 360-365.

Hahn 2002
Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *Journal of Evaluation in Clinical Practice* 2002; 8: 353-359.

Hansson 1999
Hansson L, Lindholm LH, Niskanen L, Lanke J, Hedner T, Niklason A, Luomanmaki K, Dahlof B, de Faire U, Morlin C, Karlberg BE, Wester PO, Bjorck JE. Effect of angiotensin-converting-

enzyme inhibition compared with conventional therapy on cardiovascular morbidity and mortality in hypertension: the Captopril Prevention Project (CAPPP) randomised trial. *The Lancet* 1999; 353: 611-616.

Hill 1990
Hill AB. Memories of the British streptomycin trial in tuberculosis: the first randomized clinical trial. *Controlled Clinical Trials* 1990; 11: 77-79.

Hollis 1999
Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999; 319: 670-674.

Hróbjartsson 2007
Hróbjartsson A, Forfang E, Haahr MT, Is-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *International Journal of Epidemiology* 2007; 36: 654-663.

Hutton 2000
Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Journal of the Royal Statistical Society Series C* 2000; 49: 359-370.

Jadad 1996
Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, McQuay H. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials* 1996; 17: 1-12.

Johansen 2000
Johansen HK, Gøtzsche PC. Amphotericin B lipid soluble formulations versus amphotericin B in cancer patients with neutropenia. *Cochrane Database of Systematic Reviews* 2000, Issue 3. Art No: CD000969.

Jørgensen 2006
Jørgensen KJ, Johansen HK, Gøtzsche PC. Voriconazole versus amphotericin B in cancer patients with neutropenia. *Cochrane Database of Systematic Reviews* 2006, Issue 1. Art No: CD004707.

Jørgensen 2007
Jørgensen KJ, Johansen HK, Gøtzsche PC. Flaws in design, analysis and interpretation of Pfizer's antifungal trials of voriconazole and uncritical subsequent quotations. *Trials* 2007; 7: 3.

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: 1054-1060.

Jüni 2001
Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001; 323: 42-46.

Kjaergard 2001
Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Annals of Internal Medicine* 2001; 135: 982-989.

Lachin 2000
Lachin JM. Statistical considerations in the intent-to-treat principle. *Controlled Clinical Trials* 2000; 21: 167-189.

Marshall 2000
Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *British Journal of Psychiatry* 2000; 176: 249-52.

Melander 2003
Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine - selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003; 326: 1171-1173.

Moher 1995
Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials* 1995; 16: 62-73.

Moher 1996
Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: Current issues and future directions. *International Journal of Technology Assessment in Health Care* 1996; 12: 195-208.

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? *The Lancet* 1998; 352: 609-613.

Moher 2001
Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *The Lancet* 2001; 357: 1191-1194.

Montori 2002
Montori VM, Bhandari M, Devereaux PJ, Manns BJ, Ghali WA, Guyatt GH. In the dark: the reporting of blinding status in randomized controlled trials. *Journal of Clinical Epidemiology* 2002; 55: 787-790.

Naylor 1997
Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. *BMJ* 1997; 315: 617-619.

Newell 1992
Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *International Journal of Epidemiology* 1992; 21: 837-841.

Noseworthy 1994
Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology* 1994; 44: 16-20.

Oxman 1993
Oxman AD, Guyatt GH. The science of reviewing research. *Annals of the New York Academy of Sciences* 1993; 703: 125-133.

Peto 1999
Peto R. Failure of randomisation by "sealed" envelope. *The Lancet* 1999; 354: 73.

Pildal 2007
Pildal J, Hróbjartsson A, Jørgensen KJ, Hilden J, Altman DG, Gøtzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. *International Journal of Epidemiology* 2007; 36: 847-857.

Porta 2007
Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. *Journal of Clinical Epidemiology* 2007; 60: 663-669.

Rees 2005
Rees JR, Wade TJ, Levy DA, Colford JM, Jr., Hilton JF. Changes in beliefs identify unblinding in randomized controlled trials: a method to meet CONSORT guidelines. *Contemporary Clinical Trials* 2005; 26: 25-37.

Sackett 2007
Sackett DL. Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't? *International Journal of Epidemiology* 2007; 36: 664-665.

Safer 2002
Safer DJ. Design and reporting modifications in industry-sponsored comparative psychopharmacology trials. *Journal of Nervous and Mental Disease* 2002; 190: 583-592.

Schulz 1995a
Schulz KF. Subverting randomization in controlled trials. *JAMA* 1995; 274: 1456-1458.

Schulz 1995b
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: 408-412.

Schulz 1996
Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996; 312: 742-744.

Schulz 2002a
Schulz KF, Chalmers I, Altman DG. The landscape and lexicon of blinding in randomized trials. *Annals of Internal Medicine* 2002; 136: 254-259.

Schulz 2002b
Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *The Lancet* 2002; 359: 614-618.

Schulz 2002c
Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *The Lancet* 2002; 359: 515-519.

Schulz 2002d
Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. *The Lancet* 2002; 359: 966-970.

Schulz 2006
Schulz KF, Grimes DA. *The Lancet Handbook of Essential Concepts in Clinical Research*. Edinburgh (UK): Elsevier, 2006.

Schulz 2010
Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: c332.

Senn 1991
Senn S. Baseline comparisons in randomized clinical trials. *Statistics in Medicine* 1991; 10: 1157-1159.

Siersma 2007
Siersma V, Is-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. *Statistics in Medicine* 2007; 26: 2745-2758.

Smilde 2001
Smilde TJ, van Wissen S, Wollersheim H, Trip MD, Kastelein JJ, Stalenhoef AF. Effect of aggressive versus conventional lipid lowering on atherosclerosis progression in familial hypercholesterolaemia (ASAP): a prospective, randomised, double-blind trial. *The Lancet* 2001; 357: 577-581.

Spiegelhalter 2003
Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statistics in Medicine* 2003; 22: 3687-3709.

Sterne 2002
Sterne JA, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Statistics in Medicine* 2002; 21: 1513-1524.

Tierney 2005
Tierney JF, Stewart LA. Investigating patient exclusion bias in meta-analysis. *International Journal of Epidemiology* 2005; 34: 79-87.

Turner 2008
Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society Series A* 2008 (published online ahead of print, 8 July 2008).

Unnebrink 2001
Unnebrink K, Windeler J. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. *Statistics in Medicine* 2001; 20: 3931-3946.

Vickers 2001
Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Medical Research Methodology* 2001; 1: 6.

von Elm 2006
von Elm E, Röllin A, Blümle A, Senessie C, Low N, Egger M. Selective reporting of outcomes of drug trials. Comparison of study protocols and pulbished articles. 14th Cochrane Colloquium, Dublin (Ireland), 2006.

Williamson 2005a
Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. *Statistics in Medicine* 2005; 24: 1547-1561.

Williamson 2005b
Williamson PR, Gamble C, Altman DG, Hutton JL. Outcome selection bias in meta-analysis. *Statistical Methods in Medical Research* 2005; 14: 515-524.

Wood 2004
Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials* 2004; 1: 368-376.

Wood 2008
Wood L, Egger M, Gluud LL, Schulz K, Jüni P, Altman DG, Gluud C, Martin RM, Wood AJG, Sterne JAC. *Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study*. *BMJ* 2008; 336: 601-605.

Woods 1995
Woods KL. Mega-trials and management of acute myocardial infarction. *The Lancet* 1995; 346: 611-614.

ÍNDICE
CAPÍTULO 9:
ANÁLISIS DE LOS DATOS Y REALIZACIÓN DE LOS METANÁLISIS

PUNTOS CLAVE

9.1 INTRODUCCIÓN

9.1.1 ¡No comience aquí!
9.1.2 Planificación del análisis
9.1.3 ¿Por qué realizar un metanálisis en una revisión?
9.1.4 Cuándo no utilizar un metanálisis en una revisión
9.1.5 ¿Qué implica un metanálisis?
9.1.6 ¿Qué comparaciones se deben hacer?
9.1.7 Redacción de la sección de análisis del protocolo

9.2 TIPO DE DATOS Y MEDIDAS DE EFECTO

9.2.1 Tipo de datos
9.2.2 Medidas del efecto para los resultados dicotómicos
Recuadro 9.2.a: Cálculo del razón de riesgos (RR), el odds ratio (OR) y la diferencia de riesgos (DR) de una tabla de 2×2
    9.2.2.1 Riesgo y odds
    9.2.2.2 Medidas de efecto relativo: la razón de riesgos y el odds ratio
    9.2.2.3 Advertencia: OR y CR no son lo mismo
    9.2.2.4 Medida de efecto absoluto: la diferencia de riesgos
    9.2.2.5 ¿Qué es el suceso?
9.2.3 Medidas de efecto para los resultados continuos
    9.2.3.1 La diferencia de medias (o diferencia en las medias)
    9.2.3.2 La diferencia de medias estandarizada
9.2.4 Medidas de efecto para resultados ordinales y escalas de medición
9.2.5 Medidas de efecto para recuentos y tasas
    9.2.5.1 Advertencia: ¿Recuento de sucesos o recuento de participantes?
9.2.6 Medidas de efecto para resultados de tiempo hasta el suceso (supervivencia)
9.2.7 Expresión de los efectos de la intervención en escalas logarítmicas

9.3 DISEÑOS DE LOS ESTUDIOS E IDENTIFICACIÓN DE LA UNIDAD DE ANÁLISIS

9.3.1 Problemas relacionados con la unidad de análisis
9.3.2 Ensayos aleatorizados por conglomerados
9.3.4 Observaciones repetidas en los participantes
9.3.5 Sucesos que pueden volver a ocurrir
9.3.6 Múltiples intentos de tratamiento
9.3.7 Múltiples partes del cuerpo I: las partes del cuerpo reciben la misma intervención
9.3.8 Múltiples partes del cuerpo II: las partes del cuerpo reciben intervenciones diferentes
9.3.9 Múltiples grupos de intervención

9.4 RESUMEN DE LOS EFECTOS ENTRE LOS ESTUDIOS

9.4.1 Metanálisis
9.4.2 Principios del metanálisis
9.4.3 Enfoque genérico del inverso de la varianza para el metanálisis
    9.4.3.1 Método de efectos aleatorios (DerSimonian y Laird) para el metanálisis
    9.4.3.2 El tipo de resultado inverso de la varianza genérico en RevMan
9.4.4 Metanálisis de resultados dicotómicos

9.4.4.1 Métodos de Mantel-Haenszel
9.4.4.2 Método del odds ratio de Peto
9.4.4.3 Método de efectos aleatorios
9.4.4.4 ¿Qué medida utilizar para los resultados dicotómicos?
9.4.5 Metanálisis de resultados continuos
9.4.5.1 ¿Qué medida utilizar para los resultados continuos?
9.4.5.2 Metanálisis de las puntuaciones de cambio
9.4.5.3 Metanálisis de datos asimétricos
9.4.6 Combinación de resultados dicotómicos y continuos
9.4.7 Metanálisis de resultados ordinales y escalas de medición
9.4.8 Metanálisis de recuentos y tasas
9.4.9 Metanálisis de resultados de tiempo hasta el suceso
9.4.10 Un resumen de los métodos de metanálisis disponibles en RevMan
Tabla 9.4.a Resumen de los métodos de metanálisis disponibles en RevMan
9.4.11 Uso del recuento de votos para el metanálisis

9.5 HETEROGENEIDAD

9.5.1 ¿Qué es heterogeneidad?
9.5.2 Identificación y medición de la heterogeneidad
9.5.3 Estrategias para abordar la heterogeneidad
9.5.4 Incorporación de la heterogeneidad en los modelos de efectos aleatorios

9.6 INVESTIGACIÓN DE LA HETEROGENEIDAD

9.6.1 Interacción y modificación del efecto
9.6.2 ¿Qué son los análisis de subgrupos?
9.6.3 Realización de los análisis de subgrupos
9.6.3.1 ¿El efecto es diferente en subgrupos diferentes?
9.6.4 Metarregresión
9.6.5 Selección de las características del estudio para los análisis de subgrupos y la metarregresión
9.6.5.1 Asegúrese de que haya estudios adecuados para justificar los análisis de subgrupos y las metarregresiones
9.6.5.2 Especifique de antemano las características
9.6.5.3 Seleccione un pequeño número de características
9.6.5.4 Asegúrese de que hay una justificación científica para investigar cada característica
9.6.5.5 Tenga en cuenta que no siempre es posible identificar el efecto de una característica
9.6.5.6 Considere si la característica está muy relacionada con otra característica (factor de confusión)
9.6.6 Interpretación de los análisis de subgrupos y las metarregresiones
9.6.7 Investigación del efecto del riesgo basal
9.6.8 Análisis de dosis respuesta

9.7 ANÁLISIS DE SENSIBILIDAD

9.8 INFORMACIÓN DEL CAPÍTULO

Recuadro 9.8.a: El Grupo Cochrane de Métodos Estadísticos

9.9 REFERENCIAS

CAPÍTULO 9: ANÁLISIS DE LOS DATOS Y REALIZACIÓN DE LOS METANÁLISIS

Editores: Jonathan J Deeks, Julian PT Higgins y Douglas G Altman del Grupo Cochrane de Métodos Estadísticos.

Puntos clave

- El metanálisis es la combinación estadística de los resultados de dos o más estudios individuales.
- Las ventajas potenciales del metanálisis incluyen un aumento de la potencia estadística, una mejoría en la precisión, la capacidad de responder preguntas no planteadas en los estudios individuales y la oportunidad de resolver controversias que surgen de conclusiones contradictorias. Sin embargo, también pueden provocar confusiones importantes, particularmente si no se analizan cuidadosamente los diseños específicos de los estudios, los sesgos dentro de los estudios, las variaciones entre los estudios y los sesgos de informe.
- Es importante estar familiarizado con el tipo de datos (p.ej. dicotómicos, continuos) que surge de la medición de un resultado en un estudio individual, y seleccionar las medidas de efecto adecuadas para comparar los grupos de intervención.
- La mayoría de los métodos para el metanálisis son variaciones de un promedio ponderado de las estimaciones del efecto de diferentes estudios.
- Se deben tomar en cuenta las variaciones entre los estudios (heterogeneidad), aunque la mayoría de las revisiones Cochrane no tienen estudios suficientes para permitir la investigación confiable de los motivos para dicha heterogeneidad. Los metanálisis de efectos aleatorios permiten la presencia de heterogeneidad al asumir que los efectos subyacentes siguen una distribución normal.
- En el proceso de preparar una revisión Cochrane o un metanálisis es necesario tomar muchas decisiones. Se deben utilizar los análisis de sensibilidad para examinar si los hallazgos generales son robustos frente a decisiones potencialmente influyentes.

9.1 Introducción

9.1.1 ¡No comience aquí!

Puede ser tentador pasar prematuramente al análisis estadístico cuando se realiza una revisión sistemática. La producción de un diamante al final de un gráfico es un momento excitante para muchos revisores, pero los resultados de los metanálisis pueden dar lugar a muchos errores si no se presta suficiente atención a formular la pregunta de la revisión; especificar los criterios de elegibilidad; identificar, seleccionar y evaluar críticamente los estudios; obtener los datos apropiados y decidir que será importante analizar. Los revisores deben consultar los capítulos anteriores antes de realizar un metanálisis.

9.1.2 Planificación del análisis

Mientras que en los estudios primarios los investigadores seleccionan y obtienen los datos de pacientes individuales, en las revisiones sistemáticas los investigadores seleccionan y obtienen los datos de los estudios primarios. Mientras que los estudios primarios incluyen los análisis de sus participantes, las revisiones Cochrane contienen análisis de los estudios primarios. Los análisis pueden ser narrativos, como un resumen estructurado y una discusión de las características y hallazgos de los estudios, o cuantitativos e incluir análisis estadísticos. El metanálisis (la combinación estadística de los resultados de dos o más estudios individuales) es la técnica estadística más utilizada. El programa informático para desarrollar una revisión Cochrane (RevMan) puede realizar una variedad de metanálisis, pero se debe señalar que el metanálisis no es apropiado para todas las revisiones Cochrane. Los aspectos a considerar cuando se decide si un metanálisis es apropiado en una revisión se discuten en esta sección y en la Sección 9.1.4.

Los estudios que comparan intervenciones de atención sanitaria, principalmente ensayos aleatorizados, utilizan los resultados de los participantes para comparar los efectos de diferentes intervenciones. Los metanálisis se centran en comparaciones apareadas de las intervenciones, como una intervención experimental versus una intervención control, o la comparación de dos intervenciones experimentales. La terminología utilizada aquí (intervención experimental versus control) incluye el primer ejemplo, aunque los métodos se aplican de forma similar al segundo.

El contraste entre los resultados de dos grupos tratados de forma diferente se conoce como el “efecto”, el “efecto del tratamiento” o el “efecto de la intervención”. Si el análisis de los estudios incluidos es narrativo o cuantitativo, se debe proporcionar un marco general para la síntesis que considere los cuatro interrogantes siguientes:

1. ¿Cuál es la dirección del efecto?
2. ¿Cuál es el tamaño del efecto?
3. ¿El efecto es consistente entre los estudios?
4. ¿Cuál es la fuerza de la evidencia del efecto?

El metanálisis proporciona un método estadístico para los interrogantes 1 a 3. La evaluación del interrogante 4 considera adicionalmente valoraciones basadas en las evaluaciones del diseño del estudio y el riesgo de sesgo, así como en las medidas estadísticas de incertidumbre.

La síntesis narrativa utiliza métodos subjetivos (en lugar de estadísticos) para responder los interrogantes 1 a 4, en revisiones donde el metanálisis no es posible ni sensible. En una síntesis narrativa se debe preespecificar, justificar y seguir sistemáticamente el método utilizado para cada estadio. Es posible introducir sesgo si se hace un énfasis inapropiado en los resultados de un estudio en detrimento de los de otro.

El plan de análisis se deriva del objetivo científico de la revisión. Las revisiones pueden tener diferentes tipos de objetivos, por lo que también pueden tener diferentes enfoques para el análisis.

1. La revisión Cochrane más sencilla combina estudios que hacen una comparación particular entre dos opciones de tratamiento, por ejemplo, comparación de extracto de kava versus placebo para tratar la ansiedad (Pittler 2003). Se pueden utilizar el metanálisis y las técnicas relacionadas si existe una medida de resultado consistente para:
- Establecer si existe prueba de un efecto;
- Realizar la estimación del tamaño del efecto y la incertidumbre alrededor de dicho tamaño; e
- Investigar si el efecto es consistente entre los estudios.

2. Algunas revisiones pueden tener un foco más amplio que una comparación única. El primer caso es cuando la intención es identificar y cotejar estudios de numerosas intervenciones para la misma enfermedad o afección. Un ejemplo de este tipo de revisión es la de los tratamientos tópicos para las infecciones micóticas de la piel y las uñas de los pies, que incluye estudios de cualquier tratamiento tópico (Crawford 2007). En el segundo caso el objetivo relacionado es identificar la "mejor" intervención. Una revisión de intervenciones para la anticoncepción de emergencia investiga cuál es la más efectiva (y al mismo tiempo considera los efectos adversos potenciales). Dichas revisiones pueden incluir múltiples comparaciones y metanálisis entre todos los pares posibles de tratamientos, y se requiere precaución cuando se van a planificar los análisis (ver Sección 9.1.6 y Capítulo 16, Sección 16.6).

3. En ocasiones las comparaciones de la revisión tienen espectros particularmente amplios que hacen que el uso del metanálisis sea problemático. Por ejemplo, una revisión de intervenciones en el puesto de trabajo para el abandono del hábito de fumar incluyó diferentes tipos de intervenciones (Moher 2005). Cuando las revisiones contienen estudios muy diversos un metanálisis pudiera ser útil para responder la pregunta general de si existen pruebas de que, por ejemplo, las intervenciones en el puesto de trabajo pueden funcionar (no obstante, ver Sección 9.1.4). Pero es posible que el uso del metanálisis para describir el tamaño del efecto no sea significativo si las implementaciones son tan diferentes que no sea posible interpretar una estimación del efecto en algún contexto específico.

4. Un objetivo de algunas revisiones es investigar la relación entre el tamaño de un efecto y alguna/s característica/s de los estudios. Lo anterior no es frecuente como objetivo primario de una revisión Cochrane pero puede ser un objetivo secundario. Por ejemplo, en una revisión de beclometasona versus placebo para el asma crónica, hubo interés en si la dosis de beclometasona administrada afectaba la eficacia (Adams 2005). Estas investigaciones de la heterogeneidad se deben realizar con precaución (ver Sección 9.6).

9.1.3 ¿Por qué realizar un metanálisis en una revisión?

El valor que un metanálisis puede aportar a una revisión depende del contexto en el cual se utiliza, como se describe en la Sección 9.1.2. Los siguientes son motivos para considerar incluir un metanálisis en una revisión.
1. Aumentar la potencia estadística. La potencia estadística es la probabilidad de detectar un efecto real estadísticamente significativo, si el mismo existe. Muchos estudios individuales son demasiado pequeños para detectar efectos pequeños, pero cuando se combinan varios de ellos hay una mayor probabilidad de detectar un efecto.
2. Aumentar la precisión. La estimación del efecto de una intervención puede mejorar cuando se basa en más información.
3. Responder preguntas no planteadas en los estudios individuales. A menudo los estudios primarios incluyen un tipo específico de pacientes e intervenciones explícitamente definidas. Una selección de estudios en los cuales estas características difieran puede permitir la

investigación de la consistencia del efecto y, si es relevante, permitir investigar los motivos de las diferencias en las estimaciones del efecto.
4. Resolver controversias que surgen de estudios aparentemente contradictorios o generar nuevas hipótesis. El análisis estadístico de los hallazgos permite evaluar formalmente el grado de conflicto, así como explorar y cuantificar los motivos de los resultados diferentes.

Por supuesto, el uso de métodos estadísticos no garantiza que los resultados de una revisión sean más validos que los del estudio primario. Además, como cualquier herramienta, los métodos estadísticos pueden utilizarse de forma inapropiada.

9.1.4 Cuándo no utilizar un metanálisis en una revisión

Si se utiliza de forma apropiada, el metanálisis es una herramienta poderosa para derivar conclusiones significativas de los datos y puede ayudar a evitar errores en la interpretación. Sin embargo, existen situaciones en las cuales el metanálisis puede ser más un obstáculo que una ayuda.

- Una crítica frecuente a los metanálisis es que “combinan manzanas con peras”. Si los estudios son clínicamente diferentes entonces un metanálisis puede carecer de sentido y enmascarar las diferencias reales en los efectos. Un tipo particularmente importante de diversidad son las comparaciones que se realizan en los estudios primarios. A menudo no tienen sentido combinar todos los estudios incluidos en un único metanálisis, algunas veces existe una mezcla de comparaciones de tratamientos diferentes con comparadores diferentes, y es posible que cada combinación de ellos se deba considerar por separado. Además, es importante no combinar resultados que sean demasiado diversos. Las decisiones acerca de qué se debe o no combinar son inevitablemente subjetivas y no son susceptibles de solución estadística, sino que requieren discusión y valoración clínica. En algunos casos puede ser difícil alcanzar el consenso.
- Los metanálisis de los estudios con riesgo de sesgo pueden llevar a conclusiones muy erróneas. Si existe sesgo en todos (o algunos) de los estudios individuales, el metanálisis simplemente combinará los errores y producirá un resultado “erróneo”, el cual se puede interpretar como de mayor credibilidad.
- Finalmente, es probable que los metanálisis en presencia de sesgos graves de publicación o informe generen un resumen inapropiado.

9.1.5 ¿Qué implica un metanálisis?

Aunque el uso de métodos estadísticos en las revisiones puede ser extremadamente útil, el elemento fundamental de un análisis es un enfoque detallado de sus elementos narrativos y cuantitativos. Lo anterior implica considerar las siguientes preguntas:
1. ¿Qué comparaciones se deben hacer?
2. ¿Qué resultados del estudio se deben utilizar en cada comparación?
3. ¿Cuál es el mejor resumen del efecto para cada comparación?
4. ¿Los resultados de los estudios son similares dentro de cada comparación?
5. ¿Cuán fiables son estos resúmenes?

El primer paso para abordar estas preguntas es decidir qué comparaciones hacer (ver Sección 9.1.6) y qué tipo de datos son apropiados para el resultado de interés (ver Sección 9.2). El paso siguiente es preparar resúmenes tabulados de las características y los resultados de los estudios incluidos en cada comparación (la extracción de los datos y su conversión a un formato conveniente se discuten en el Capítulo 7, Sección 7.7). Es posible entonces derivar de una manera sistemática las estimaciones del efecto entre los estudios (Sección 9.4), para medir e investigar las diferencias entre los estudios (Secciones 9.5 y 9.6) e interpretar los hallazgos y concluir cuánta confianza se puede tener en los mismos (ver Capítulo 12).

9.1.6 ¿Qué comparaciones se deben hacer?

El primer paso, y el más importante, al planificar el análisis es especificar las comparaciones apareadas que se harán. Las comparaciones abordadas en la revisión se deben relacionar clara y directamente con las preguntas o hipótesis que se plantean cuando se formula la revisión (ver Capítulo 5). Debería ser posible especificar en el protocolo de una revisión las comparaciones principales que se harán. Sin embargo, a menudo será necesario modificar las comparaciones y agregar nuevas a partir de los datos que se obtienen. Por ejemplo, cuando variaciones importantes en la intervención sólo se descubren después de obtener los datos.

Las decisiones acerca de qué estudios son suficientemente similares para agrupar sus resultados requiere comprensión del problema que aborda la revisión, así como de valoraciones del revisor y el usuario. La formulación de las preguntas que aborda una revisión se discute en el Capítulo 5. Esencialmente, las mismas consideraciones se aplican a la decisión de qué comparaciones hacer, qué resultados combinar y qué características clave (del diseño del estudio, los participantes, las intervenciones y los resultados) considerar cuando se investigan las variaciones del efecto (heterogeneidad). Estas consideraciones se deben abordar cuando se configuran las tablas “Datos y análisis” en RevMan y cuando se decide qué información colocar en la tabla “Características de los estudios incluidos”.

9.1.7 Redacción de la sección de análisis del protocolo

La sección “Análisis” del protocolo de una revisión Cochrane puede ser más susceptible de cambiar que otras secciones del protocolo (como los criterios para incluir los estudios o cómo se evaluará la calidad metodológica). Pocas veces es posible anticipar todos los problemas estadísticos que pueden surgir, por ejemplo, el hallazgo de resultados que son similares pero no son los mismos entre sí, resultados medidos en puntos temporales múltiples o variados y el uso de tratamientos concomitantes.

Sin embargo, el protocolo debe proporcionar una indicación firme sobre cómo el revisor enfocará la evaluación estadística de los hallazgos de los estudios. Al menos un miembro del equipo de revisión debe estar familiarizado con la mayoría de los contenidos de este capítulo cuando se redacte el protocolo. Como una guía se recomienda que se aborden los siguientes aspectos:

1. Asegurar que la estrategia de análisis aborde firmemente los objetivos señalados en la revisión (ver Sección 9.1.2).
2. Considerar qué tipo de diseño de estudios sería el apropiado para la revisión. Los ensayos de grupos paralelos son lo habitual, pero otros diseños aleatorios pueden ser apropiados para el tema (p.ej. ensayos cruzados, ensayos aleatorizados por conglomerados, ensayos factoriales). Decidir cómo se abordarán estos estudios en el análisis (ver Sección 9.3).
3. Decidir si se planifica realizar un metanálisis y considerar cómo se tomará la decisión de si es apropiado realizar un metanálisis (ver Secciones 9.1.3 y 9.1.4).
4. Determinar la naturaleza probable de los datos de resultado (p.ej. dicotómicos, continuos, etc.) (ver Sección 9.2).
5. Considerar si es posible especificar por adelantado qué medidas del efecto de la intervención se utilizarán (p.ej. razón de riesgos, odds ratios o diferencia de riesgos para los resultados dicotómicos; diferencia de medias o diferencia de medias estandarizada para los resultados continuos) (ver Secciones 9.4.4.4 y 9.4.5.1).
6. Decidir cómo se identificará o cuantificará la heterogeneidad estadística (ver Sección 9.5.2).
7. Decidir si se utilizarán metanálisis de efectos aleatorios, metanálisis de efectos fijos o ambos métodos para cada metanálisis planificado (ver Sección 9.5.4).
8. Considerar cómo se evaluará la diversidad clínica y metodológica (heterogeneidad) y si la misma se incorporará en la estrategia de análisis (y cómo) (ver Secciones 9.5 y 9.6).
9. Decidir cómo se evaluará e incluirá en el análisis el riesgo de sesgo de los estudios incluidos (ver Capítulo 8).

10. Preespecificar las características de los estudios que se pueden examinar como posibles causas de heterogeneidad (ver Sección 9.6.5).
11. Considerar cómo se tratarán los datos faltantes (p.ej. imputación de los datos o análisis por intención de tratar) (ver Capítulo 16, Secciones 16.1 y 16.2).
12. Decidir si se explorarán las pruebas de posibles sesgos de publicación o informe (y cómo) (ver Capítulo 10).

Durante la redacción de un protocolo puede resultar evidente la necesidad de conocimientos adicionales y, de ser así, se debe buscar un estadístico para que se una al grupo de revisión.

9.2 Tipo de datos y medidas de efecto

9.2.1 Tipo de datos

El punto de partida de todos los metanálisis de los estudios de efectividad incluye la identificación del tipo de datos para las medidas de resultado. En este capítulo se considera que los datos de resultado son de cinco tipos diferentes:
1. datos dicotómicos (o binarios), donde cada resultado de un individuo es uno de sólo dos respuestas categóricas posibles;
2. datos continuos, donde cada resultado de un individuo es una medición de una cantidad numérica;
3. datos ordinales (incluidas las escalas de medición), donde el resultado es uno de varias categorías ordenadas, o se genera al sumar las respuestas categóricas de una escala;
4. recuentos y tasas calculadas del recuento del número de sucesos que presenta cada individuo; y
5. datos de tiempo hasta el suceso (habitualmente supervivencia) que analizan el tiempo hasta que ocurre un suceso, pero donde no todos los individuos en el estudio presentan el suceso (datos censurados).

La forma en la que se puede medir el efecto de una intervención depende de la naturaleza de los datos que se obtienen. En esta sección se examinan brevemente los tipos de datos de resultado que se pudieran encontrar en las revisiones sistemáticas de los ensayos clínicos, así como las definiciones, propiedades e interpretación de las medidas estándar del efecto de la intervención. En las secciones 9.4.4.4 y 9.4.5.1 se discuten los aspectos relacionados con la selección de una de estas medidas para un metanálisis particular.

9.2.2 Medidas del efecto para los resultados dicotómicos

Los datos de resultado dicotómicos (binarios) surgen cuando el resultado para cada participante es una de dos posibilidades, por ejemplo, muerto o vivo, o mejoría clínica o ninguna mejoría clínica. Esta sección considera los estadísticos resumen disponibles cuando el resultado de interés tiene una forma binaria. Las medidas del efecto encontradas y utilizadas con más frecuencia en los ensayos clínicos con datos dicotómicos son:
- la razón de riesgos (RR) (también llamado riesgo relativo);
- el odds ratio (OR);
- la diferencia de riesgos (DR) (también llamada reducción del riesgo absoluto); y
- el número necesario a tratar (NNT).

Los detalles de los cálculos de las tres primeras medidas se proporcionan en el Recuadro 9.2.a. Los números necesarios a tratar se discuten en detalle en el Capítulo 12 (Sección 12.5).

Aparte: Como en ocasiones los sucesos pueden ser deseables en lugar de indeseables, sería preferible utilizar un término más neutral que "riesgo" (como probabilidad), pero de forma convencional se utilizan los términos razón de riesgos y diferencia de riesgo todo el tiempo. También se utiliza el término "razón de riesgos" en lugar de "riesgo relativo" para ser consistentes con otra terminología. Los dos son intercambiables y en inglés se abrevian

convenientemente como "RR". También es de señalar que se ha tenido mucho cuidado al utilizar las palabras "riesgo" y "tasas". Estas palabras a menudo se tratan como sinónimos. Sin embargo, se ha tratado de reservar el uso de la palabra "tasa" para el tipo de datos "recuentos y tasas" donde el mismo describe la frecuencia de sucesos en un período de tiempo medido.

Recuadro 9.2.a: Cálculo del razón de riesgos (RR), el odds ratio (OR) y la diferencia de riesgos (DR) de una tabla de 2×2

Los resultados de un ensayo clínico se pueden mostrar como una tabla de 2×2:

	Suceso ("Éxito")	Ningún suceso ("Fracaso")	Total
Intervención experimental	E_E	F_E	N_E
Intervención control	E_C	F_C	N_C

donde E_E, E_C, F_E y F_C son los números de participantes con cada resultado ("E" o "F") en cada grupo ("E" o "C"). Se pueden calcular las siguientes estadísticas resumen:

$$\text{CR}=\frac{\text{riesgo de evento en grupo experimental}}{\text{riesgo de evento en grupo control}}=\frac{E_E/N_E}{E_C/N_C}$$

$$\text{OR}=\frac{\text{posibilidad de evento en grupo experimental}}{\text{posibilidad de evento en grupo control}}=\frac{E_E/F_E}{E_C/F_C}=\frac{E_E F_C}{F_E E_C}$$

$$\text{DR}=\text{riesgo de evento en grupo experimental - riesgo de evento en grupo control}=\frac{E_E}{N_E}-\frac{E_C}{N_C}$$

9.2.2.1 Riesgo y odds

En la conversación general los términos "riesgo" y "odds" se utilizan indistintamente como si los mismos describieran la misma cantidad. Sin embargo, en estadística el riesgo y el odds tienen significados particulares y se calculan de forma diferente. Cuando se ignoran las diferencias entre ellos, los resultados de las revisiones sistemáticas pueden interpretarse de forma errónea.

Riesgo es el concepto más familiar para los pacientes y los profesionales de la salud. El riesgo describe la probabilidad de que ocurra un resultado de salud (habitualmente un suceso adverso). En investigación, el riesgo se expresa frecuentemente como un número decimal entre 0 y 1, aunque en ocasiones se convierte a un porcentaje. En las tablas "Síntesis de los resultados" de las revisiones Cochrane a menudo se expresa como un número de individuos por 1000 (ver Capítulo 11, Sección 11.5). Es sencillo comprender la relación entre un riesgo y la probabilidad de que ocurra un suceso: en una muestra de 100 personas el número de sucesos observados será como promedio el riesgo multiplicado por 100. Por ejemplo, cuando el riesgo es 0,1 en cerca de 10 personas de cada 100 se observará el suceso; cuando el riesgo es 0,5 en cerca de 50 personas cada 100 se observará el suceso. En una muestra de 1000 personas, los números son 100 y 500, respectivamente.

El odds es un concepto que es más familiar para los jugadores. El odds es el cociente de la probabilidad de que un suceso particular ocurra sobre la probabilidad de que el suceso no ocurra, y puede ser un número entre 0 e infinito. En el juego, el odds describe el cociente del tamaño de las ganancias potenciales para la apuesta del juego; en la atención sanitaria es el cociente del número de personas con el suceso sobre el número de personas sin el suceso. Frecuentemente se expresa como un cociente de dos números. Por ejemplo, un odds de 0,01

a menudo se escribe como 1:100; un odds de 0,33 como 1:3 y un odds de 3 como 3:1. El odds se puede convertir a riesgos, y los riesgos a odds, mediante la fórmula:

$$\text{riesgo} = \frac{\text{posibilidad}}{1 + \text{posibilidad}}; \text{posibilidad} = \frac{\text{riesgo}}{1\text{- riesgo}}$$

La interpretación de un odds es más complicada que la de un riesgo. La forma más sencilla de asegurar que la interpretación es correcta es convertir primero un odds en un riesgo. Por ejemplo, cuando el odds es 1:10 o 0,1 el suceso se observará en una persona por cada 10 que no se observará, y utilizando la fórmula, el riesgo del suceso es 0,1/(1+0,1) = 0,091. En una muestra de 100, en cerca de 9 individuos se observará el suceso y en 91 no. Cuando el odds es igual a 1, se observará el suceso en una persona por cada una que no, por lo que en una muestra de 100, 100 × 1/(1+1) = 50 se observará el suceso y 50 no.

La diferencia entre el odds y el riesgo es pequeña cuando el suceso es poco frecuente (como se ilustra en el primer ejemplo anteriormente, donde se observó que un riesgo de 0,091 es similar a un odds de 0,1). Cuando los sucesos son frecuentes, como ocurre a menudo en el caso de los ensayos clínicos, las diferencias entre el odds y el riesgo son grandes. Por ejemplo, un riesgo de 0,5 equivale a un odds de 1; y un riesgo de 0,95 equivale a un odds de 19.

Las medidas del efecto para los ensayos clínicos con resultados dicotómicos se basan en comparar los riesgos o los odds de dos grupos de intervención. Para compararlos se puede examinar su cociente (la razón de riesgo o el odds ratio) o la diferencia en los riesgos (diferencia de riesgos).

9.2.2.2 Medidas de efecto relativo: la razón de riesgos y el odds ratio

Las medidas de efecto relativo expresan el resultado en un grupo con relación al resultado de otro. La razón de riesgos (o riesgo relativo) es el cociente del riesgo de un suceso en los dos grupos, mientras que el odds ratio es el cociente de los odds de un suceso (ver Recuadro 9.2.a). Para ambas medidas un valor de 1 indica que los efectos estimados son los mismas para ambas intervenciones.

Ni el riesgo relativo ni el odds ratio se pueden calcular en un estudio si no hay sucesos en el grupo control. Esto se debe, como se puede observar en la fórmula del Recuadro 9.2.a, a que se estaría dividiendo por cero. El odds ratio tampoco se puede calcular si todos los sujetos del grupo intervención presentan un suceso. En estas situaciones, y otras donde no es posible computar los errores estándar, es habitual agregar ½ a cada celda de la tabla de 2x2 (RevMan hace esta corrección automáticamente si es necesario). En el caso donde no se observen sucesos o se observen en todos los sujetos en ambos grupos, el estudio no proporciona información acerca de la probabilidad relativa del suceso y se omite automáticamente del metanálisis. El procedimiento anterior es completamente correcto. Los ceros aparecen particularmente cuando el suceso de interés es poco frecuente (a menudo estos sucesos son resultados adversos no planificados). Para una discusión adicional de la selección de las medidas de efecto para datos escasos (frecuentemente con muchos ceros) ver Capítulo 16 (Sección 16.9).

Las razones de riesgos describen la multiplicación del riesgo que ocurre con el uso de la intervención experimental. Por ejemplo, una razón de riesgos de 3 para un tratamiento implica que los sucesos con el tratamiento son tres veces más probables que los sucesos sin el tratamiento. Alternativamente, es posible decir que el tratamiento aumenta el riesgo de sucesos en 100 x (CR-1)% = 200%. De manera similar, una razón de riesgos de 0,25 se interpreta como que la probabilidad de que ocurra un suceso con el tratamiento es la cuarta parte de que ocurra sin el tratamiento. Lo anterior se puede expresar de forma alternativa al decir que el tratamiento disminuye el riesgo del suceso en 100 x (1-CR)% = 75%, lo cual se

conoce como reducción del riesgo relativo (ver también Capítulo 12, Sección 12.5.1). La interpretación de la importancia clínica de una razón de riesgos determinada no se puede hacer sin conocer el riesgo típico de los sucesos sin tratamiento; una razón de riesgos de 0,75 pudiera corresponder con una reducción clínicamente importante en los sucesos del 80% al 60%, o con una reducción pequeña y clínicamente menos importante del 4% al 3%.

El valor numérico de la razón de riesgos observada siempre debe estar entre 0 y 1/RGC, donde el RGC (abreviatura de "riesgo del grupo control", algunas veces referido como tasa de suceso control) es el riesgo observado del suceso en el grupo control (expresado como un número entre 0 y 1). Lo anterior significa que para sucesos frecuentes son imposibles valores grandes de la razón de riesgos. Por ejemplo, cuando el riesgo observado de sucesos en el grupo control es 0,66 (o 66%), la razón de riesgos observada no puede exceder de 1,5. Este problema se aplica solamente para los aumentos del riesgo y sólo es problemático cuando los resultados se extrapolan a riesgos por encima de los observados en el estudio.

Los odds ratios, como los odds, son más difíciles de interpretar (Sinclair 1994, Sackett 1996). Los odds ratios describen la multiplicación de los odds del desenlace que ocurre con el uso de la intervención. Para comprender qué significa un odds ratio en términos de cambios en el número de sucesos es más sencillo convertirlo primero en una razón de riesgos, y luego interpretar la razón de riesgos en el contexto de un riesgo típico del grupo control, como se esboza anteriormente. La fórmula para convertir un odds ratio en una razón de riesgos se proporciona en el Capítulo 12 (Sección 122.5.4.4). Algunas veces puede ser adecuado calcular el CR para más de un supuesto riesgo del grupo control.

9.2.2.3 Advertencia: OR y CR no son lo mismo

Debido a que el riesgo y el odds toman valores distintos cuando los sucesos son frecuentes, la razón de riesgos y el odds ratio también difieren cuando los sucesos son frecuentes. La falta de equivalencia entre la razón de riesgos y el odds ratio no indica que alguno de ellos sea erróneo; ambos son formas completamente válidas de describir el efecto de una intervención. Sin embargo, los problemas pueden surgir si el odds ratio se malinterpreta como una razón de riesgos. Para las intervenciones que aumentan las probabilidades de los sucesos, el odds ratio será mayor que la razón de riesgos, por lo que la interpretación errónea puede tender a sobrestimar el efecto de la intervención, especialmente cuando los sucesos son frecuentes (por ejemplo, con riesgos de sucesos mayores del 20%). Para intervenciones que reducen la probabilidad de los sucesos el odds ratio será menor que la razón de riesgos, por lo que nuevamente la interpretación errónea sobrestima el efecto de la intervención. Desafortunadamente este error en la interpretación es bastante frecuente en los informes publicados de los estudios individuales y las revisiones sistemáticas.

9.2.2.4 Medida de efecto absoluto: la diferencia de riesgos

La diferencia de riesgos es la diferencia entre los riesgos observados (proporciones de individuos con el resultado de interés) en los dos grupos (ver Recuadro 9.2.a). La diferencia de riesgos se puede calcular para cualquier estudio, incluso cuando no hay sucesos en los grupos. La diferencia de riesgos es sencilla de interpretar, describe la diferencia real en el riesgo observado de sucesos entre las intervenciones experimental y control; para un individuo describe la diferencia de la estimación en la probabilidad de presentar el suceso. Sin embargo, la importancia clínica de la diferencia de riesgos puede depender del riesgo del suceso subyacente. Por ejemplo, una diferencia de riesgos de 0,02 (o 2%) puede representar un cambio pequeño clínicamente no significativo de un riesgo del 58% al 60% o un cambio proporcionalmente mucho mayor y potencialmente importante del 1% al 3%. Aunque la diferencia de riesgos proporciona información con una relevancia más directa que las medidas relativas (Laupacis 1988, Sackett 1997), sigue siendo importante estar informado del riesgo de sucesos subyacente y de las consecuencias de los sucesos cuando se interpreta una diferencia de riesgos. Las medidas absolutas como la diferencia de riesgos son

particularmente útiles cuando se considera el balance entre los probables beneficios y daños de una intervención.

La diferencia de riesgos está limitada de forma natural (al igual que la razón de riesgos), lo cual puede crear dificultades cuando se aplican los resultados a otros grupos de pacientes y otros contextos. Por ejemplo, si en un estudio o un metanálisis se calcula una diferencia de riesgos de -0,1 (o -10%), entonces para un grupo con un riesgo inicial del 7%, por ejemplo, el resultado tendrá una estimación imposible de la probabilidad negativa del -3%. En el otro extremo de la escala aparecen escenarios similares para el aumento del riesgo. Estos problemas sólo pueden surgir cuando los resultados se aplican a pacientes con riesgos diferentes de los observados en los estudios.

El número necesario a tratar se obtiene a partir de la diferencia de riesgos. Aunque a menudo se utiliza para resumir los resultados de los ensayos clínicos, los NNT no se pueden combinar en un metanálisis (ver Sección 9.4.4.4). Sin embargo, los odds ratios, las razones de riesgos y las diferencias de riesgos se pueden convertir apropiadamente en NNT y utilizarse cuando se interpretan los resultados de un metanálisis, como se discute en el Capítulo 12 (Sección 12.5).

9.2.2.5 ¿Qué es el suceso?

En el contexto de los resultados dicotómicos, las intervenciones de la atención sanitaria tienen como objetivo reducir el riesgo de ocurrencia de un suceso adverso o aumentar la probabilidad de un buen resultado. Todas las medidas de efecto descritas en la Sección 9.2.2 se aplican por igual a ambos escenarios.

En muchas situaciones es natural denominar al estado del resultado como un suceso. Por ejemplo, cuando los participantes presentan síntomas particulares al inicio del estudio, habitualmente el suceso de interés es recuperación o curación. Si los participantes están bien o, alternativamente, en riesgo de algún suceso adverso al inicio del estudio, entonces el suceso es la aparición de la enfermedad o la ocurrencia del suceso adverso. Debido a que generalmente el foco se centra en el grupo de la intervención experimental, un estudio en el cual la intervención experimental reduzca la ocurrencia de un resultado tendrá un odds ratio y una razón de riesgos menores de 1, y una diferencia de riesgos negativa. Un estudio en el cual la intervención experimental aumente la ocurrencia de un resultado tendrá un odds ratio y una razón de riesgos mayores de 1, y una diferencia de riesgos positiva (ver Recuadro 9.2.a).

Sin embargo, es posible intercambiar suceso y ningún suceso y considerar en su lugar la proporción de pacientes que no se recuperaron o que no presentaron el suceso. Para los metanálisis que utilizan las diferencias de riesgos o los odds ratios la repercusión de este intercambio no tiene mayores consecuencias; el intercambio implica cambios en el signo de una diferencia de riesgos, mientras que para los odds ratios el nuevo odds ratio es el reciproco ($1/x$) del odds ratio original.

Por el contrario, el intercambio del resultado puede provocar diferencias significativas para las razones de riesgos que afectan la estimación del efecto, su significación y la consistencia de los efectos de la intervención entre los estudios. Lo anterior se debe a que la precisión de la estimación de una razón de riesgos difiere marcadamente entre las situaciones donde los riesgos son bajos y las situaciones donde los riesgos son altos. En un metanálisis el efecto de esta inversión no se puede predecir fácilmente. Por lo tanto, previo al análisis de los datos es importante identificar qué cociente de riesgos es más probable que sea el estadístico resumen más relevante, lo que se discute en la Sección 9.4.4.4.

9.2.3 Medidas de efecto para los resultados continuos

El término "continuo" en estadística se refiere convencionalmente a los datos que pueden tomar cualquier valor en un rango especificado. Cuando se trata de datos numéricos, lo anterior significa que es posible medir e informar cualquier número hasta cualquier posición decimal. Ejemplos de datos verdaderamente continuos son el peso, el área y el volumen. En la práctica, en las revisiones Cochrane se pueden utilizar los mismos métodos estadísticos para otros tipos de datos, con mayor frecuencia las escalas de medición y los recuentos de grandes números de sucesos (ver Sección 9.2.4).

Hay dos estadísticos resumen que se utilizan con frecuencia para el metanálisis de los datos continuos: la diferencia de medias y la diferencia de medias estandarizada. Las mismas se pueden calcular tanto si los datos de cada individuo son evaluaciones únicas como si son medidas de cambio respecto los valores iniciales. También es posible medir los efectos a través de los cocientes de las medias, o al comparar estadísticos que no sean las medias (p.ej. las medianas). Sin embargo, los métodos para éstos no se abordan aquí.

9.2.3.1 La diferencia de medias (o diferencia en las medias)

La diferencia de medias (y no media de las diferencias) es un estadístico estándar que mide la diferencia absoluta entre el valor medio en dos grupos de un ensayo clínico. La misma calcula la magnitud en la cual la intervención experimental cambia el resultado como promedio, comparada con el control. Se puede utilizar como un estadístico resumen en el metanálisis cuando las mediciones del resultado en todos los estudios se realizan con la mima escala.

Aparte: A los análisis basados en esta medida de efecto se les ha denominado históricamente diferencia de medias ponderada (DMP) en la Base de Datos Cochrane de Revisiones Sistemáticas *(BDCRS). Este nombre es potencialmente confuso: aunque el metanálisis calcula un promedio ponderado de estas diferencias de medias, en el cálculo de un estadístico resumen de un estudio único no se incluye una ponderación. Además, todos los metanálisis incluyen una combinación ponderada de las estimaciones, pero no se utiliza la palabra "ponderado" para referirse a otros métodos.*

9.2.3.2 La diferencia de medias estandarizada

La diferencia de medias estandarizada se utiliza como un estadístico resumen en el metanálisis cuando todos los estudios evalúan el mismo resultado pero lo miden de varias formas (por ejemplo, todos los estudios miden depresión pero utilizan diferentes escalas psicométricas). En este caso es necesario estandarizar los resultados de los estudios en una escala uniforme antes de combinarlos. La diferencia de medias estandarizada expresa el tamaño del efecto de la intervención en cada estudio con relación a la variabilidad observada en ese estudio. (Nuevamente, en realidad el efecto de la intervención es una diferencia de medias y no una media de las diferencias.):

$$\text{DME} = \frac{\text{diferencia de medias de resultados entre grupos}}{\text{desviación estándar de resultados de participantes}}.$$

Por lo tanto, los estudios para los cuales la diferencia de medias es la misma proporción de la desviación estándar tendrán la misma DME, independientemente de las escalas utilizadas para realizar las mediciones.

Sin embargo, el método supone que las diferencias en las desviaciones estándar entre los estudios reflejan diferencias en las escalas de medición y no diferencias reales en la variabilidad entre las poblaciones de los estudios. Esta suposición puede ser problemática en algunas circunstancias donde se esperan diferencias verdaderas en la variabilidad entre los

participantes de diferentes estudios. Por ejemplo, cuando se combinan ensayos pragmáticos y explicatorios en la misma revisión los ensayos pragmáticos pueden incluir un rango mayor de participantes, por lo que pueden tener mayores desviaciones estándar. También puede ser difícil de interpretar el efecto general de la intervención cuando el mismo se informa en unidades de desviación estándar en lugar en unidades de cualquiera de las escalas de medición utilizadas en la revisión, pero en algunas circunstancias es posible transformar el efecto nuevamente a las unidades utilizadas en el estudio específico (ver Capítulo 12, Sección 12.6).

El término "tamaño del efecto" (*effect size*) se utiliza frecuentemente en las ciencias sociales, particularmente en el contexto del metanálisis. Habitualmente, pero no siempre, el tamaño del efecto se refiere a versiones de la diferencia de medias estandarizada. Se recomienda que en las revisiones Cochrane se utilice el término "diferencia de medias estandarizada", en lugar de "tamaño del efecto", para evitar confusiones con el uso médico más general del último término como sinónimo de "efecto de la intervención" o "estimación del efecto". La definición particular de la diferencia de medias estandarizada utilizada en las revisiones Cochrane es el tamaño del efecto conocido en ciencias sociales como *g* de Hedges (ajustada).

Se debe señalar que el método de la DME no corrige las diferencias en la dirección de la escala. Si algunas escalas aumentan con la gravedad de la enfermedad mientras que otras disminuyen, es fundamental multiplicar los valores medios de un grupo de estudios por -1 (o alternativamente, restar la media del valor máximo posible para la escala) para asegurar que todas las escalas apunten en la misma dirección. Cualquier ajuste de este tipo se debe describir en la sección "Métodos estadísticos" de la revisión. No es necesario modificar la desviación estándar.

9.2.4 Medidas de efecto para resultados ordinales y escalas de medición

Los datos de resultado ordinales surgen cuando cada participante se clasifica en una categoría y cuando las categorías tienen un orden natural. Por ejemplo, un resultado "tricotómico" con un ordenamiento de las categorías como la clasificación de la gravedad de la enfermedad en "leve", "moderada" o "grave", es de tipo ordinal. Cuando el número de categorías aumenta, los resultados ordinales adquieren propiedades similares a los resultados continuos y probablemente deberán analizarse como tales en un ensayo clínico.

Las escalas de medición son un tipo particular de resultado ordinal utilizado frecuentemente para medir afecciones difíciles de cuantificar como la conducta, la depresión y las capacidades cognitivas. Las escalas de medición generalmente incluyen una serie de preguntas o tareas, cada una de las cuales recibe una puntuación y las puntuaciones se suman para producir una "puntuación total". Si se considera que los ítems no tienen una importancia similar se puede utilizar una suma ponderada.

Es importante conocer si las escalas han sido validadas, es decir, si se ha probado que miden las condiciones que dicen medir. Cuando una escala se utiliza para evaluar un resultado en un ensayo clínico, se debe estudiar la referencia citada para la escala con el fin de entender el objetivo, la población a quien está dirigida y el cuestionario de evaluación. Como frecuentemente los investigadores adaptan las escalas para que se ajusten a su propio objetivo al agregar, cambiar u obviar preguntas, los revisores deben verificar si se utilizó un cuestionario original o adaptado. Lo anterior es particularmente importante cuando los resultados se agrupan para un metanálisis. Puede parecer que los ensayos clínicos utilizan la misma escala de medición, pero un examen más cuidadoso muestra diferencias que se deben tomar en cuenta. Es posible que las modificaciones a una escala se hagan debido a los resultados de un estudio, con el fin de destacar componentes de una intervención experimental que parecen ser beneficiosos.

Existen métodos especializados para analizar los datos de resultado ordinales que describen los efectos en términos de odds ratios proporcionales, pero los mismos no están

disponibles en RevMan, y se hacen difíciles de manejar (e innecesarios) cuando el número de categorías es grande. En la práctica las escalas ordinales largas a menudo se analizan en los metanálisis como datos continuos, mientras que las escalas ordinales más cortas se convierten frecuentemente en datos dicotómicos al combinar juntas las categorías adyacentes. Este último procedimiento es especialmente apropiado cuando está disponible un punto de corte establecido y defendible. La selección inapropiada de un punto de corte puede inducir sesgo, particularmente si el mismo se selecciona para maximizar las diferencias entre dos brazos de intervención en un ensayo clínico.

Cuando las escalas ordinales se resumen mediante los métodos utilizados para los datos dicotómicos, uno de los dos conjuntos de categorías agrupadas se define como el suceso y los efectos de la intervención se describen mediante las razones de riesgos, los odds ratios o las diferencias de riesgos (ver Sección 9.2.2). Cuando las escalas ordinales se resumen mediante los métodos para los datos continuos, el efecto de la intervención se expresa como una diferencia de medias o una diferenta de medias estandarizada (ver Sección 9.2.3). Se encontrarán dificultades si los estudios resumieron sus resultados mediante medianas (ver Capítulo 7, Sección 7.7.3.5).

A menos que los datos de pacientes individuales estén disponibles, los análisis informados por los investigadores en los ensayos clínicos determinan el enfoque a utilizar en el metanálisis.

9.2.5 Medidas de efecto para recuentos y tasas

Algunos tipos de sucesos pueden ocurrir más de una vez en una persona, por ejemplo, un infarto de miocardio, una fractura, reacciones adversas o ingresos hospitalarios. Pudiera ser recomendable o necesario abordar el número de veces que ocurren estos sucesos en lugar de determinar sencillamente sin cada persona presentó un suceso (es decir, en lugar de tratarlos como datos dicotómicos). A este tipo de datos se les denomina datos de recuento. Para fines prácticos los datos de recuento se pueden dividir convenientemente en recuentos de sucesos poco frecuentes y recuentos de sucesos frecuentes.

Frecuentemente a los sucesos de recuentos se les denomina en estadística “datos de Poisson”. El análisis de los sucesos poco frecuentes a menudo se basa en las tasas. Las tasas relacionan los recuentos con la cantidad de tiempo durante el cual pudieron haber ocurrido. Por ejemplo, el resultado de un brazo de un ensayo clínico pudiera ser que se presentaron 18 infartos de miocardio (IM) entre los participantes en ese brazo durante un período de 314 personas-año de seguimiento. La tasa es 0,057 por persona-año o 5,7 por 100 personas-año. El estadístico resumen utilizado habitualmente en el metanálisis es el cociente de tasas (también abreviado como CT), que compara la tasa de sucesos en los dos grupos al dividir uno por el otro. También es posible utilizar una diferencia de tasas como estadístico resumen, aunque es mucho menos frecuente.

Los recuentos de sucesos más frecuentes, como los recuentos de los dientes cariados, perdidos y obturados, a menudo se pueden tratar de la misma manera que los datos de resultado continuos. El efecto de la intervención utilizado será la diferencia de medias, la cual comparará la diferencia en el número medio de sucesos (posiblemente estandarizados según una unidad de tiempo) presentados por los participantes del grupo intervención, en comparación con los participantes del grupo control.

9.2.5.1 Advertencia: ¿Recuento de sucesos o recuento de participantes?

Un error frecuente es intentar tratar los datos de recuentos como datos dicotómicos. Suponga que en el ejemplo recién presentado las 314 personas-año surgen de 157 participantes observados durante, como promedio, dos años. Es tentador citar los resultados como 18/157. Lo anterior es inapropiado si múltiples IM del mismo paciente pueden haber contribuido al total de 18 (por ejemplo, si los 18 ocurrieron en 12 pacientes que tuvieron IM únicos y 3 pacientes presentaron 2 IM cada uno). El número total de sucesos pudiera teóricamente

exceder el número de pacientes, lo que hace que los resultados no tengan sentido. Por ejemplo, en el curso de un año, 35 pacientes con epilepsia en un estudio presentaron 63 convulsiones entre ellos.

9.2.6 Medidas de efecto para resultados de tiempo hasta el suceso (supervivencia)

Los datos de tiempo hasta el suceso surgen cuando el interés se centra en el tiempo que transcurre antes de que ocurra un suceso. En estadística se conocen genéricamente como datos de supervivencia, ya que a menudo la muerte es el resultado de interés, particularmente en el cáncer y las cardiopatías. Los datos de tiempo hasta el suceso consisten en pares de observaciones para cada individuo: (i) un período de tiempo durante el cual no se observan sucesos, y (ii) un indicador de si el final de ese período de tiempo corresponde con un suceso o es sólo el final de la observación. Los participantes que contribuyen con algún período de tiempo que no termina en un suceso se les lama “censurados”. Su tiempo sin suceso contribuye con información y se incluyen en el análisis. Los datos de tiempo hasta el suceso se pueden basar en suceso que no sean la muerte, como recurrencia de un suceso de la enfermedad (por ejemplo, tiempo hasta el final de un período sin crisis epilépticas) o hasta el alta del hospital.

En ocasiones los datos de tiempo hasta el suceso se pueden analizar como datos dicotómicos. Lo anterior requiere conocer el estado de todos los pacientes en un estudio, en un punto temporal fijo. Por ejemplo, si todos los pacientes se han seguido durante al menos 12 meses, y la proporción de los que han presentado el suceso antes de los 12 meses es conocida para ambos grupos, entonces es posible construir una tabla de 2x2 (ver Recuadro 9.2.a) y expresar los efectos de la intervención como razones de riesgos, odds ratios y diferencias de riesgos.

No es apropiado analizar los datos de tiempo hasta el suceso mediante métodos para resultados continuos (p.ej. la media de los tiempos hasta el suceso), ya que los tiempos relevantes se conocen solamente par el subgrupo de participantes que han presentado el suceso. Se deben excluir los pacientes censurados, lo cual es casi seguro que introducirá sesgo.

La forma más apropiada de resumir los datos de tiempo hasta el suceso es utilizar métodos de análisis de supervivencia y expresar el efecto de la intervención como un cociente de riesgos instantáneos (*hazard ratio*). La noción del cociente de riesgos instantáneos es similar a la del riesgo, pero difiere en que mide el riesgo instantáneo y puede cambiar constantemente (por ejemplo, su propio cociente de riesgos instantáneos de muerte cambia cuando usted cruza una carretera muy transitada). Un cociente de riesgos instantáneos se interpreta de una manera similar al cociente de riesgos, ya que describe cuántas veces más (o menos) probable es que un participante presente un suceso en un punto temporal particular si reciben la intervención experimental en lugar de la intervención control. Cuando se comparan las intervenciones en un estudio o un metanálisis, a menudo se hace una suposición simple de que el cociente de riesgos instantáneos es constante durante todo el período de seguimiento, incluso aunque los cocientes de riesgos instantáneos por sí mismos pueden variar constantemente. Lo anterior se conoce como la hipótesis de cocientes de riesgos instantáneos proporcionales.

9.2.7 Expresión de los efectos de la intervención en escalas logarítmicas

A los valores del cociente de los efectos de la intervención (como odds ratio, cociente de riesgos, cociente de tasas y cociente de riesgos instantáneos) habitualmente se les realizan transformaciones logarítmicas antes de analizarlos, y ocasionalmente se les puede llamar en términos de sus valores transformados logarítmicamente. Generalmente se utiliza la transformación logarítmica *natural* (log base *e*, que se escribe “ln”).

Todos los estadísticos resumen basados en cocientes tienen en común que el valor más bajo pueden alcanzar es 0, que el valor 1 corresponde con ningún efecto de la intervención, y que el valor más alto que puede tener un odds ratio es infinito. Esta escala numérica no es simétrica. Por ejemplo, mientras un odds ratio de 0,5 (la mitad) y un OR de 2 (el doble) son opuestos, por lo que deberían promediar un efecto nulo, el promedio de 0,5 y 2 no es un OR de 1 sino de 1,25. La transformación logarítmica hace que la escala sea simétrica: el logaritmo de 0 es menos infinito, el logaritmo de 1 es cero y el logaritmo de infinito es infinito. En el ejemplo, el logaritmo del OR de 0,5 es -0,69 y el logaritmo del OR de 2 es 0,69. El promedio de -0,69 y 0,69 es 0, que es el valor transformado logarítmicamente de un OR de 1, lo que implica de forma correcta que no hay un efecto promedio de la intervención.

Habitualmente la presentación gráfica de un metanálisis realizado con escalas de cocientes utiliza una escala logarítmica. Lo anterior tiene el efecto de hacer que los intervalos de confianza parezcan simétricos, por los mismos motivos.

9.3 Diseños de los estudios e identificación de la unidad de análisis

9.3.1 Problemas relacionados con la unidad de análisis

Un principio importante en los ensayos clínicos es que el análisis debe tener en cuenta el nivel al cual se realizó la asignación al azar. En la mayoría de las circunstancias el número de observaciones en el análisis debe coincidir con el número de “unidades” que se asignaron al azar. En un diseño simple de grupos paralelos para un ensayo clínico, los participantes se asignan al azar individualmente a uno de dos grupos de intervención, y se obtiene y analiza una medición única para cada resultado de cada participante. Sin embargo, existen numerosas variaciones de este diseño. Los revisores deben considerar si en cada estudio:

- grupos de individuos se asignaron al azar a una misma intervención (es decir, ensayos aleatorizados por conglomerados);
- los individuos recibieron más de una intervención (p.ej. en un ensayo cruzado, o tratamientos simultáneos en múltiples partes de cada individuo); y
- existen varias observaciones para el mismo resultado (p.ej. mediciones repetidas, sucesos recurrentes, mediciones en diferentes partes del cuerpo).

A continuación se presenta una lista más detallada de situaciones en las cuales surgen frecuentemente problemas con la unidad de análisis, junto con indicaciones a discusiones relevantes en alguna de las secciones del *Manual*.

9.3.2 Ensayos aleatorizados por conglomerados

En un ensayo aleatorizado por conglomerados los grupos de participantes se asignan al azar a diferentes intervenciones. Por ejemplo, los grupos pueden ser escuelas, pueblos, prácticas médicas, pacientes de un único doctor o familias. Ver Capítulo 16 (Sección 16.3).

9.3.3 Ensayos cruzados

En un ensayo cruzado todos los participantes reciben todas las intervenciones en secuencia: se asignan a un orden de las intervenciones y los participantes actúan como su propio control. Ver Capítulo 16 (Sección 16.4).

9.3.4 Observaciones repetidas en los participantes

En los estudios de larga duración los resultados se pueden presentar para varios períodos de seguimiento (por ejemplo, a los seis meses, un año o dos años). Los resultados de más de un punto temporal para cada estudio no se pueden combinar en un metanálisis estándar sin que surja un error en la unidad de análisis. Algunas opciones son las siguientes:

- Obtener los datos de pacientes individuales y realizar un análisis (como el análisis del tiempo hasta el suceso) que utilice el seguimiento total para cada participante. Alternativamente, computar una medida de efecto para cada participante individual que incorpore todos los puntos temporales, como el número total de sucesos, una media general o una tendencia en el tiempo. Ocasionalmente dichos análisis están disponibles en los informes publicados.
- Definir varios resultados diferentes basados en diferentes períodos de seguimiento y realizar análisis separados. Por ejemplo, se pudieran definir marcos de tiempo que reflejen el seguimiento a corto, medio y largo plazo.
- Seleccionar un punto temporal único y analizar solamente los datos de este momento para los estudios en los cuales se presenta. Idealmente debe ser un punto temporal clínicamente importante. Algunas veces este procedimiento se puede seleccionar para maximizar los datos disponibles, aunque los revisores deben estar al tanto de la posibilidad de sesgo de informe.
- Seleccionar el seguimiento más largo de cada estudio. Lo anterior puede inducir una falta de consistencia entre los estudios y dar lugar a heterogeneidad.

9.3.5 Sucesos que pueden volver a ocurrir

Si el resultado de interés es un suceso que puede ocurrir más de una vez, entonces se debe tener precaución para evitar un error de la unidad de análisis. Los datos de recuentos no se deben tratar como si fueran datos dicotómicos. Ver Sección 9.2.5.

9.3.6 Múltiples intentos de tratamiento

De manera similar, los múltiples intentos de tratamiento por participante pueden provocar un error de la unidad de análisis. Se debe tener la precaución de asegurar que se utilice el número de participantes asignados al azar y no el número de intentos de tratamiento para calcular los intervalos de confianza. Por ejemplo, en estudios de subfertilidad a las mujeres se les pueden realizar múltiples ciclos, y los revisores pudieran utilizar erróneamente como denominador los ciclos en lugar de las mujeres. Esta situación es similar en los ensayos aleatorizados por conglomerados, excepto que cada participante es el “grupo”. Ver los métodos descritos en el Capítulo 16 (Sección 16.3).

9.3.7 Múltiples partes del cuerpo I: las partes del cuerpo reciben la misma intervención

En algunos estudios las personas se asignan al azar pero múltiples partes (o lugares) del cuerpo reciben la misma intervención, se realiza una valoración separada del resultado para cada parte del cuerpo y el número de partes del cuerpo se utiliza como denominador en el análisis. Por ejemplo, los ojos se pueden utilizar erróneamente como el denominador, sin ajustar por la falta de independencia entre ellos. Esta situación es similar en los ensayos aleatorios por conglomerados, excepto que cada participante es el “grupo”. Ver los métodos descritos en el Capítulo 16 (Sección 16.3).

9.3.8 Múltiples partes del cuerpo II: las partes del cuerpo reciben intervenciones diferentes

Una situación diferente es aquella donde las diferentes partes del cuerpo se asignan al azar a intervenciones *diferentes*. Los diseños “Boca dividida” en salud bucal son de este tipo, en el cual diferentes áreas de la boca se asignan a intervenciones diferentes. Estos ensayos tienen semejanzas con los ensayos cruzados; mientras que en los ensayos cruzados los individuos reciben tratamientos múltiples en momentos diferentes, en estos ensayos reciben tratamientos múltiples en diferentes sitios. Ver los métodos descritos en el Capítulo 16 (Sección 16.4). Es importante distinguir estos estudios de aquellos donde los participantes reciben la misma intervención en múltiples sitios (Sección 9.3.7).

9.3.9 Múltiples grupos de intervención

Los estudios que comparan más de dos grupos de intervención se deben tratar con precaución. A menudo estos estudios se incluyen en el metanálisis haciendo comparaciones pareadas múltiples entre todos los pares posibles de grupos de intervención. Surge un problema serio con la unidad de análisis si el mismo grupo de participantes se incluye dos veces en el mismo metanálisis (por ejemplo, si "Dosis 1 versus Placebo" y "Dosis 2 versus Placebo" se incluyen en el mismo metanálisis, con los mismos pacientes placebo en cada comparación). Ver Capítulo 16 (Sección 16.5).

9.4 Resumen de los efectos entre los estudios

9.4.1 Metanálisis

Un paso importante en una revisión sistemática es la consideración cuidadosa de si es apropiado combinar los resultados numéricos de algunos o todos los estudios. Dicho metanálisis produce un estadístico general (junto con su intervalo de confianza) que resume la efectividad de la intervención experimental comparada con una intervención control (ver Sección 9.1.2). Esta sección describe los principios y métodos utilizados para realizar un metanálisis de los principales tipos de datos encontrados.

Las fórmulas para todos los métodos descritos se proporcionan en un documento complementario, Algoritmos estadísticos en Review Manager 5 (disponible en el sitio web del *Manual*), y en Deeks y cols. (Deeks 2001) aparece una discusión más larga de los aspectos analizados en esta sección.

9.4.2 Principios del metanálisis

Todos los métodos utilizados de forma habitual para el metanálisis cumplen los siguientes principios básicos:

1. El metanálisis es típicamente un proceso de dos estadios. En el primer estadio se calcula un estadístico resumen para cada estudio para describir el efecto de la intervención observado. Por ejemplo, el estadístico resumen puede ser una razón de riesgos si los datos son dicotómicos o una diferencia de medias si los datos son continuos.
2. En el segundo estadio se calcula una estimación resumen (combinada) del efecto de la intervención como un promedio ponderado de las estimaciones de los efectos de la intervención en los estudios individuales. Un promedio ponderado se define como

$$\text{promedio ponderado} = \frac{\text{suma de (cálculo x ponderación)}}{\text{suma de ponderaciones}} = \frac{\sum Y_i W_i}{\sum W_i}$$

donde Y_i es la estimación del efecto de la intervención en el estudio *i*; W_i es la ponderación dada al estudio *i*, y la suma se realiza entre todos los estudios. Es de señalar que si todas las ponderaciones son las mismas, el promedio ponderado es igual al efecto medio de la intervención. Mientras mayor sea la ponderación dada al estudio *i*, más contribuirá el mismo al promedio ponderado. Por lo tanto, las ponderaciones se seleccionan para reflejar la cantidad de información que contiene cada estudio. Para las medidas de cocientes (OR, CR, etc.) Y_i es el logaritmo natural de la medida.

3. La combinación de las estimaciones del efecto de la intervención entre los estudios puede incorporar opcionalmente una hipótesis de que no todos los estudios estiman el mismo efecto de la intervención, sino que estiman efectos de la intervención que siguen una distribución entre los estudios. Lo anterior es la base del metanálisis de efectos aleatorios (ver Sección 9.5.4.). Alternativamente, si se supone que cada estudio estima exactamente la misma cantidad, se realiza un metanálisis de efectos fijos.

4. El error estándar del efecto resumen (combinado) de la intervención se puede utilizar para derivar un intervalo de confianza, el cual informa la precisión (o incertidumbre) de la estimación resumen, y para derivar un valor de p, el cual informa la solidez de las pruebas contra la hipótesis nula de ningún efecto de la intervención.
5. De la misma manera que producen una cuantificación resumen del efecto combinado, todos los métodos de metanálisis pueden incorporar una evaluación de si la variación entre los resultados de los estudios es compatible con la variación aleatoria, o si la misma es suficientemente grande para indicar una inconsistencia de los efectos de la intervención entre los estudios (ver Sección 9.5).

9.4.3 Enfoque genérico del inverso de la varianza para el metanálisis

Existe una versión frecuente y sencilla del procedimiento del metanálisis a la que se le llama el método del inverso de la varianza. Este enfoque està implementado en RevMan en su modalidad más básica y se utiliza entre bastidores en ciertos metanálisis de datos dicotómicos y continuos.

Al método del inverso de la varianza se le llama así porque la ponderación dada a cada estudio se selecciona de manera que sea el inverso de la varianza de la estimación del efecto (es decir, uno sobre el cuadrado de su error estándar). Por lo tanto, a los estudios más grandes que tienen errores estándar más pequeños se les da una mayor ponderación que a los estudios más pequeños que tienen errores estándar mayores. Esta selección de la ponderación minimiza la imprecisión (incertidumbre) de la estimación del efecto combinada.

Un metanálisis de efectos fijos que utiliza el método del inverso de la varianza calcula un promedio ponderado como

$$\text{media ponderada de varianza inversa genérica} = \frac{\sum Y_i (1/EE_i^2)}{\sum (1/EE_i^2)},$$

donde Y_i es la estimación del efecto de la intervención en el estudio *i*, EE_i es el error estándar de dicha estimación y la suma se realiza entre todos los estudios. Por lo tanto, los datos básicos requeridos para el análisis son una estimación del efecto de la intervención y el error estándar de cada estudio.

9.4.3.1 Método de efectos aleatorios (DerSimonian y Laird) para el metanálisis

Una variación del método del inverso de la varianza es incorporar una hipótesis de que los diferentes estudios estiman efectos de la intervención diferentes, aunque relacionados. Lo anterior produce un metanálisis de efectos aleatorios, y la versión más sencilla se conoce como el método de DerSimonian y Laird (DerSimonian 1986). El metanálisis de efectos aleatorios se discute en la Sección 9.5.4. Para realizar un metanálisis de efectos aleatorios se ajustan los errores estándar de las estimaciones específicas del estudio (EE_i anteriormente) para incorporar una medida del grado de variabilidad, o heterogeneidad, entre los efectos de la intervención observados en diferentes estudios (a menudo a esta variación se le llama tau cuadrado [τ^2, o Tau^2]). La cantidad de variación, y por lo tanto el ajuste, se pueden calcular a partir de los efectos de la intervención y de los errores estándar de los estudios incluidos en el metanálisis.

9.4.3.2 El tipo de resultado inverso de la varianza genérico en RevMan

Las estimaciones y sus errores estándar se pueden entrar directamente en RevMan bajo el resultado “Inverso de la varianza genérico”. El programa informático realizará los metanálisis de efectos fijos y efectos aleatorios (DerSimonian y Laird), junto con las evaluaciones de la heterogeneidad. Para las medidas de cocientes del efecto de la intervención, los datos se deben ingresar como logaritmos naturales (por ejemplo, como un logaritmo del odds ratio y el

error estándar del logaritmo del odds ratio). Sin embargo, es sencillo indicarle al programa que muestre los resultados en la escala original (p.ej. odds ratio). En lugar de mostrar los datos resumen de forma separada para los grupos de tratamiento, el diagrama de árbol mostrará las estimaciones y los errores estándar como se grabaron al lado de los identificadores del estudio. Es posible complementarlos o reemplazarlos con una columna que proporcione los tamaños de muestra en los dos grupos.

Cabe destacar que la posibilidad de entrar las estimaciones y los errores estándar directamente en RevMan permite un alto grado de flexibilidad en el metanálisis. Por ejemplo, facilita el análisis de ensayos cruzados, ensayos aleatorizados por conglomerados y estudios no aleatorizados analizados adecuadamente, así como los datos de resultado ordinales, de tiempo hasta el suceso o tasas. Sin embargo, en la mayoría de las situaciones para los análisis de los datos de resultado continuos y dicotómicos es preferible entrar datos más detallados en RevMan (es decir, específicamente como resúmenes simples de datos dicotómicos o continuos para cada grupo). Lo anterior evita la necesidad de que el revisor calcule las estimaciones del efecto y permite el uso de métodos dirigidos específicamente a diferentes tipos de datos (ver Secciones 9.4.4 y 9.4.5). Además, es útil para los lectores de la revisión ver los estadísticos resumen para cada grupo de intervención en cada estudio.

9.4.4 Metanálisis de resultados dicotómicos

Hay cuatro métodos ampliamente utilizados para el metanálisis de los resultados dicotómicos, tres métodos de efectos fijos (Mantel-Haenszel, Peto y el inverso de la varianza) y un método de efectos aleatorios (DerSimonian y Laird). Todos estos métodos están disponibles como opciones de análisis en RevMan. El método de Peto sólo puede combinar odds ratios, mientras que los otros tres métodos pueden combinar los odds ratios, las razones de riesgos y las diferencias de riesgos. Las fórmulas para todos los métodos del metanálisis se proporcionan en Deeks y cols. (Deeks 2001).

Nótese que las celdas vacías (p.ej. ningún suceso en un grupo) provocan problemas con el cálculo de las estimaciones y los errores estándar con algunos métodos. El programa informático RevMan agrega automáticamente 0,5 a cada celda de la tabla de 2x2 para cada estudio de este tipo.

9.4.4.1 Métodos de Mantel-Haenszel

Los métodos de Mantel-Haenszel (Mantel 1959, Greenland 1985) son los métodos de metanálisis de efectos fijos establecidos por defecto en RevMan. Cuando los datos son escasos, en términos de las tasas de sucesos bajas o del tamaño pequeño del estudio, el cálculo de los errores estándar de las estimaciones del efecto que se utilizan en los métodos del inverso de la varianza puede ser deficiente. Los métodos de Mantel-Haenszel utilizan un esquema de ponderación diferente que depende de qué medida de efecto (p.ej., cociente de riesgos, odds ratio, diferencia de riesgo) se utilice. Han mostrado tener mejores propiedades estadísticas cuando hay pocos sucesos. Como ésta es una situación frecuente en las revisiones Cochrane, generalmente el método de Mantel-Haenszel es preferible al método del inverso de la varianza. En otras situaciones los dos métodos dan estimaciones similares.

9.4.4.2 Método del odds ratio de Peto

El método de Peto (Yusuf 1985) sólo se puede utilizar para combinar odds ratios. Utiliza un enfoque del inverso de la varianza pero usa un método aproximado para la estimación del logaritmo del odds ratio, y aplica ponderaciones diferentes. Una forma alternativa de ver el método de Peto es como una suma de los estadísticos “O-E”. Aquí, O es el número de sucesos observado y E es el número de sucesos esperado en el grupo de intervención experimental de cada estudio.

La aproximación utilizada en el cálculo del logaritmo del odds ratio funciona bien cuando los efectos de la intervención son pequeños (los odds ratios están cerca de 1), los sucesos no son particularmente frecuentes y los estudios tienen números similares en los grupos experimental y control. En otras situaciones se ha mostrado que proporciona respuestas sesgadas. Como estos criterios no siempre se cumplen, el método de Peto no se recomienda como el enfoque por defecto para el metanálisis.

Las correcciones para las celdas con recuentos de cero no son necesarias cuando se utiliza el método de Peto. Quizás por esta razón este método funciona bien cuando los sucesos son muy poco frecuentes (Bradburn 2007) (ver Capítulo 16, Sección 16.9). Además, el método de Peto se puede utilizar para combinar estudios con datos de resultados dicotómicos y estudios que utilizan análisis de tiempo hasta el suceso cuando se han utilizado las pruebas log-rank (ver Sección 9.4.9).

9.4.4.3 Método de efectos aleatorios

El método de efectos aleatorios (DerSimonian 1986) incorpora una hipótesis de que diferentes estudios realizan la estimación de efectos de la intervención diferentes, aunque relacionados. Como se describe en la Sección 9.4.3.1, el método se basa en el enfoque del inverso de la varianza y hace un ajuste de la ponderación del estudio según el grado de variación, o heterogeneidad, entre los diferentes efectos de la intervención. El método de efectos aleatorios y el método de efectos fijos proporcionan resultados idénticos cuando no hay heterogeneidad entre los estudios. Cuando existe heterogeneidad, los intervalos de confianza para el efecto promedio de la intervención serán más amplios si se utiliza el método de efectos aleatorios en lugar del método de efectos fijos, y la significación estadística correspondiente será más conservadora. También es posible que la estimación central del efecto de la intervención cambie si existe relación entre los efectos de la intervención observados y los tamaños de muestra. Ver Sección 9.5.4. para discusión en profundidad de estos aspectos.

RevMan implementa dos métodos de efectos aleatorios para los datos dicotómicos: un método de Mantel-Haenzel y un método del inverso de la varianza. La diferencia entre ambos es sutil: el primero realiza la estimación de la cantidad de variabilidad entre estudios al comparar el resultado de cada estudio con un resultado de un metanálisis de efectos fijos de Mantel-Haenszel, mientras que el último realiza la estimación de la cantidad de variación entre estudios al comparar el resultado de cada estudio con un resultado del metanálisis de efectos fijos del inverso de la varianza. Es probable que en la práctica la diferencia sea mínima. El método del inverso de la varianza se agregó a la versión 5 de RevMan.

9.4.4.4 ¿Qué medida utilizar para los resultados dicotómicos?

Los estadísticos resumen para los datos dicotómicos se describen en la Sección 9.2.2. El efecto de la intervención se puede expresar como un efecto relativo o absoluto. La razón de riesgos (riesgo relativo) y el odds ratio son medidas relativas, y la diferencia de riesgos y el número necesario a tratar son medidas absolutas. Una complicación adicional es que de hecho existen dos razónes de riesgos. Es posible calcular la razón de riesgos de un suceso que ocurre o la razón de riesgos de un suceso que no ocurre. Ambos proporcionan resultados diferentes en el metanálisis, a veces de forma bastante marcada.

La selección de un estadístico resumen para su uso en el metanálisis depende del equilibrio entre tres criterios (Deeks 2002). Primero, es conveniente un estadístico resumen que proporcione valores que sean similares para todos los estudios en el metanálisis y subdivisiones de la población para la cual se aplicarán las intervenciones. Mientras más consistente sea el estadístico resumen mayor será la justificación para expresar el efecto de la intervención como un único número resumen. Segundo, el estadístico resumen debe tener las propiedades matemáticas necesarias para realizar un metanálisis válido. Tercero, el estadístico resumen debe ser fácilmente comprendido y aplicado por quienes utilizan la

revisión. Se debe presentar un resumen del efecto de la intervención de una manera que ayude a los lectores a interpretar y aplicar los resultados de manera apropiada. Entre las medidas de efecto para los datos dicotómicos ninguna única medida se considera de forma uniforme la mejor, por lo que la selección incluye inevitablemente un compromiso.

Consistencia: Las pruebas empíricas indican que las medidas de efecto relativas son, como promedio, más consistentes que las medidas absolutas (Engels 2000, Deeks 2002). Por este motivo es adecuado evitar realizar metanálisis de las diferencias de riesgos, a menos que exista un motivo claro para sospechar que las diferencias de riesgos serán consistentes en una situación clínica particular. En general hay muy poca diferencia entre el odds ratio y la razón de riesgos en términos de consistencia (Deeks 2002). Cuando el estudio tiene como objetivo reducir la incidencia de un resultado adverso (ver Sección 9.2.2.5), existen pruebas empíricas de que las razones de riesgos de los resultados adversos son más consistentes que las razones de riesgos de la ausencia de suceso (Deeks 2002). En general no se recomienda seleccionar una medida de efecto sobre la base de cuál es la más consistente en una situación *particular*, ya que puede dar lugar a una selección que maximice de manera espuria la precisión de una estimación del metanálisis.

Propiedades matemáticas: El criterio matemático más importante es la disponibilidad de una estimación fiable de la varianza. El número necesario a tratar no tiene un estimador único de la varianza y no es fácil utilizarlo directamente en un metanálisis, aunque se puede calcular a partir de otros estadísticos resumen (ver Capítulo 12, Sección 12.5). No hay consenso acerca de la importancia de otras dos propiedades matemáticas citadas frecuentemente: el hecho de que el comportamiento del odds ratio y la diferencia de riesgos no depende de cuál de los dos resultados señalados se codifica como el suceso, y de que el odds ratio es el único estadístico que no tiene límites (ver Sección 9.2.2).

Facilidad de interpretación: El odds ratio es el estadístico resumen más difícil de comprender y aplicar en la práctica, y muchos médicos en activo declaran tener dificultades para utilizarlo. Existen muchos ejemplos publicados donde los revisores han interpretado de manera errónea los odds ratios de los metanálisis como si se tratara de razones de riesgos. Ha habido cierta preocupación por el hecho de que la presentación habitual de los resultados de las revisiones sistemáticas como odds ratios dará lugar con frecuencia a la sobrestimación de los beneficios y daños de los tratamientos cuando los resultados se apliquen en la práctica clínica. También se considera que las medidas de efecto absolutas son más fáciles de interpretar por los médicos que los efectos relativos (Sinclair 1994), y que permiten equilibrar los posibles beneficios y daños de las intervenciones. Sin embargo, es menos probable que sean generalizables.

Parece importante evitar el uso de los estadísticos resumen para las cuales existen pruebas empíricas de que es poco probable que proporcionen estimaciones consistentes de los efectos de la intervención (la diferencia de riesgos) y no es posible utilizar estadísticos para las cuales no se puede realizar el metanálisis (el número necesario a tratar). Por lo tanto, generalmente se recomienda que para realizar el análisis se utilicen las razones de riesgos (con la precaución de realizar una selección adecuada de qué categoría de resultado se clasifica como el suceso) o los odds ratios. Puede ser conveniente realizar un análisis de sensibilidad para investigar si la selección de el estadístico resumen (y de la categoría suceso) es crítica para las conclusiones del metanálisis (ver Sección 9.7).

A menudo es apropiado utilizar un estadístico para el metanálisis y expresar nuevamente los resultados mediante un segundo estadístico más fácil de interpretar. Por ejemplo, para realizar el metanálisis con frecuencia es mejor utilizar medidas de efecto relativas (razones de riesgos u odds ratios) y expresar nuevamente el resultado a través de medidas de efecto absolutas (diferencias de riesgos o números necesarios a tratar, ver Capítulo 12, Sección 12.5). Ésta es una de las motivaciones clave para las tablas “Síntesis de los resultados” de las revisiones Cochrane (ver Capítulo 11, Sección 11.5). Si se utilizan los odds ratios para el metanálisis los mismos también se pueden expresar nuevamente como razones de riesgos

(ver Capítulo 12, Sección 12.5.4). En todos los casos se puede utilizar la misma fórmula para convertir los límites de confianza superior e inferior. Sin embargo, es importante señalar que todas estas transformaciones deben especificar un valor de riesgo inicial que indique el riesgo probable del resultado en la población "control" a la cual se le aplicará la intervención experimental. Cuando el valor seleccionado para este supuesto riesgo control es cercano a los riesgos típicos observados en el grupo control entre los estudios, se obtendrán estimaciones similares del efecto absoluto independientemente de si se utilizaron los odds ratios o los razones de riesgos para el metanálisis. Cuando el supuesto riesgo control difiere del riesgo típico observado en el grupo control, las predicciones de beneficio absoluto diferirán de acuerdo a qué estadístico resumen se utilizó para el metanálisis.

9.4.5 Metanálisis de resultados continuos

En RevMan existen dos métodos de análisis disponibles para el metanálisis de los datos continuos: el método de efectos fijos del inverso de la varianza y el método de efectos aleatorios del inverso de la varianza. Los métodos proporcionarán exactamente las mismas respuestas cuando no existe heterogeneidad. Cuando hay heterogeneidad los intervalos de confianza para un efecto promedio de la intervención serán más amplios si se utiliza el método de efectos aleatorios en lugar del método de efectos fijos, y los valores de p correspondientes serán menos significativos. También es posible que la estimación central del efecto de la intervención cambie si hay relación entre los efectos de la intervención observados y el tamaño de muestra. Ver Sección 9.5.4. para discusión adicional de estos aspectos.

Los revisores deben ser conscientes de que una hipótesis subyacente para los métodos de metanálisis de datos continuos es que los resultados siguen una distribución normal en cada brazo de intervención de cada estudio. Es posible que esta suposición no siempre se cumpla, aunque no tiene una gran importancia en estudios muy grandes. Es útil considerar la posibilidad de que existan datos asimétricos (ver Sección 9.4.5.3).

9.4.5.1 ¿Qué medida utilizar para los resultados continuos?

Existen dos estadísticos resumen utilizados para el metanálisis de los datos continuos: la diferencia de medias (DM) y la diferencia de medias estandarizadas (DME) (ver Sección 9.2.3). La selección de los estadísticos resumen para los datos continuos está determinada principalmente por el hecho de si todos los estudios informan el resultado mediante la misma escala (y entonces se puede utilizar la diferencia de medias) o mediante escalas diferentes (y entonces se debe utilizar la diferencia de medias estandarizada).

Se debe comprender los distintos roles que juegan las desviaciones estándar observadas en los dos grupos según el enfoque en la elección de estadístico resumen.

- Para el enfoque de la diferencia de medias, las desviaciones estándar se utilizan junto con el tamaño de muestra para calcular la ponderación dada a cada estudio. A los estudios con desviaciones estándar pequeñas se les da una ponderación relativamente mayor, mientras que a los estudios con desviaciones estándar grandes se les dan ponderaciones relativamente menores. Lo anterior es apropiado si la variación de las desviaciones estándar entre los estudios refleja las diferencias en la fiabilidad de las mediciones del resultado, pero es probable que no sea apropiado si las diferencias en la desviación estándar refleja diferencias reales en la variabilidad de los resultados en las poblaciones de los estudios.
- Para el enfoque de la diferencia de medias estandarizada, las desviaciones estándar se utilizan para estandarizar las diferencias de medias en una escala única (ver Sección 9.2.3.2), así como en el cálculo de las ponderaciones del estudio. Se supone que la variación entre estudios en las desviaciones estándar refleja solamente las diferencias en las escalas de medición y no las diferencias en la fiabilidad de las mediciones de resultado o la variabilidad entre las poblaciones de los estudios.

Estas limitaciones de los métodos se deben tener en mente cuando se observe una variación inesperada de las desviaciones estándar entre los estudios.

9.4.5.2 Metanálisis de las puntuaciones de cambio

En algunas circunstancias un análisis basado en los cambios respecto el valor inicial será más eficiente y tendrá una mayor potencia estadística que la comparación de los valores finales, ya que elimina del análisis el componente de la variabilidad entre las personas. Sin embargo, el cálculo de una puntuación de cambio requiere la medición del resultado en dos ocasiones y en la práctica puede ser menos eficiente para resultados inestables o que son difíciles de medir con precisión, donde el error de medición puede ser mayor que la verdadera variabilidad inicial entre las personas. Los resultados de cambio respecto el valor inicial también pueden ser más convenientes si tienen una distribución menos asimétrica que los resultados de la medición final. Aunque algunas veces se utilizan como una herramienta para "corregir" asignaciones aleatorias desafortunadas, no se recomienda esta práctica.

El enfoque estadístico más recomendado para considerar las mediciones iniciales de la variable de resultado es incluir estas mediciones iniciales del resultado como una covariable en un modelo de regresión o análisis de covarianza (ANCOVA). Estos análisis producen una estimación "ajustada" del efecto del tratamiento junto con su error estándar. Son los que se encuentran con menos frecuencia, pero como proporcionan las estimaciones más precisas y menos sesgadas de los efectos del tratamiento, los mismos se deben incluir en el análisis cuando estén disponibles. Sin embargo, sólo se pueden incluir en un metanálisis cuando se utilice el método genérico del inverso de la varianza, ya que las medias y las desviaciones estándar no están disponibles para cada grupo de intervención por separado.

En la práctica es probable que un revisor descubra que los estudios de una revisión pueden incluir una mezcla de puntuaciones de cambio respecto el valor inicial y de valores finales. Sin embargo, la mezcla de estos resultados no es un problema cuando se trata de un metanálisis de diferencias de medias. No hay motivos estadísticos para no combinar en un metanálisis estudios de resultados de cambio respecto el valor inicial con estudios de mediciones del resultado final cuando se utiliza el método de la diferencia de medias (no estandarizada) en RevMan. En un ensayo aleatorizado, habitualmente es posible asumir que las diferencias de medias basadas en cambios respecto el valor inicial abordan exactamente los mismos efectos subyacentes de la intervención que los análisis basados en las mediciones finales. Es decir, la diferencia en los valores medios finales será como promedio la misma que la diferencia en las puntuaciones de cambio medias. Si el uso de las puntuaciones de cambio aumenta la precisión, a los estudios que presentan las puntuaciones de cambio se les dará apropiadamente una ponderación mayor en el análisis que la que hubieran recibido si se hubieran utilizado los valores finales, ya que tienen menores desviaciones estándar.

Cuando se combinan los datos los revisores deben ser cuidadosos y utilizar las medias y desviaciones estándar apropiadas (ya sea de las mediciones finales o de los cambios a partir del valor inicial) para cada estudio. Como los valores de la media y las desviaciones estándar para los dos tipos de resultado pueden diferir significativamente, es recomendable colocarlas en subgrupos separados para evitar confusión a los lectores, pero es válido agrupar los resultados de los subgrupos.

Sin embargo, el valor final y las puntuaciones de cambio no se deben combinar juntas como diferencias de medias estandarizadas, ya que las diferencias en la desviación estándar no reflejan las diferencias en la escala de medición, sino en la fiabilidad de las mediciones.

Un problema práctico frecuente asociado con la inclusión de medidas de cambio respecto el valor inicial es que no se informa la desviación estándar de los cambios. Las imputaciones de las desviaciones estándar se analizan en el Capítulo 16 (Sección 16.1.3).

9.4.5.3 Metanálisis de datos asimétricos

Los análisis basados en las medias son apropiados para los datos que tienen una distribución aproximadamente normal y para los datos de ensayos muy grandes. Si la verdadera distribución de los resultados es asimétrica, se dice que los mismos son asimétricos. Algunas veces es posible detectar la asimetría a partir de las medias y las desviaciones estándar de los resultados. Es posible realizar una verificación grosera, pero solo es válida si existe un valor límite máximo o mínimo para un resultado. Por lo tanto, la verificación se puede utilizar para resultados como el peso, el volumen y las concentraciones sanguíneas, que tienen un límite inferior de 0, o para resultados de escalas con puntuaciones mínimas o máximas, pero es posible que no sea apropiada para medidas de cambio respecto el valor inicial. La verificación consiste en el cálculo de la media observada menos el valor límite inferir (o el valor límite superior menos la media observada), y dividirlo por la desviación estándar. Un cociente menor de 2 sugiere asimetría (Altman 1996). Si el cociente es menor de 1 existen pruebas sólidas de que la distribución es asimétrica.

La transformación de los datos de resultado originales puede reducir significativamente la asimetría. Los informes de ensayos pueden presentar resultados en una escala transformada, habitualmente una escala logarítmica. El enfoque recomendado en la actualidad es la obtención de resúmenes apropiados de los datos de los autores de los ensayos, o la obtención de los datos de pacientes individuales. Los resúmenes apropiados de datos y las estrategias de análisis para los datos de pacientes individuales dependerán de la situación. Se recomienda consultar a un estadístico experimentado.

Cuando los datos se analizan en una escala logarítmica, habitualmente los resultados se presentan como medias geométricas y cocientes de medias geométricas. Entonces es posible realizar un metanálisis con la escala de los datos transformados logarítmicamente. Un ejemplo del cálculo de las medias y las desviaciones estándar necesarias se brinda en el Capítulo 7 (Sección 7.7.3.4). Este enfoque depende de la posibilidad de obtener los datos transformados de todos los estudios; hay métodos disponibles para la transformación de una escala en otra (Higgins 2008b). Los datos transformados logarítmicamente y no transformados no se pueden mezclar en un metanálisis.

9.4.6 Combinación de resultados dicotómicos y continuos

En ocasiones los revisores encuentran una situación donde algunos estudios presentan los datos para el mismo resultado como dicotómicos y otros los presentan como continuos. Por ejemplo, las puntuaciones de depresión se pueden informar como medias o como el porcentaje de pacientes que presentaban depresión en algún momento después de una intervención (es decir, con una puntuación por debajo de un punto de corte especificado). A menudo este tipo de información es más fácil de comprender y más útil cuando la misma se dicotomiza. Sin embargo, la decisión de establecer un punto de corte puede ser arbitraria y se pierde información cuando los datos continuos se transforman en dicotómicos.

Existen numerosas opciones para tratar las combinaciones de datos dicotómicos y continuos. Generalmente es útil resumir los resultados de todos los estudios válidos relevantes de una forma similar, pero no siempre es posible. En ocasiones es posible obtener los datos faltantes de los investigadores para realizar dicho resumen. De no ser así, puede ser útil resumir los datos de tres maneras: registrar las medias y desviaciones estándar como resultados continuos; registrar los recuentos como resultados dicotómicos y registrar todos los datos en forma de texto como resultados “Otros datos”.

Existen enfoques estadísticos disponibles que reformularán los odds ratios como diferencias de medias estandarizadas (y *viceversa*), lo que permite agrupar juntos los datos dicotómicos y continuos. Según la suposición de que las mediciones continuas subyacentes de cada grupo de intervención siguen una distribución logística (la cual es una distribución simétrica similar en forma a la distribución normal pero con más datos en las colas de la distribución), y que la

variabilidad de los resultados es la misma en los participantes tratados y control, los odds ratios se pueden reformular como una diferencia de medias estandarizada según la simple fórmula fórmula (Chinn 2000):

$$\mathrm{SMD} = \frac{\sqrt{3}}{\pi} \ln \mathrm{OR} .$$

El error estándar del logaritmo del odds ratio se puede convertir en error estándar de la diferencia de medias estandarizada al multiplicar por la misma constante ($\sqrt{3}/\pi = 0{,}5513$). Alternativamente las diferencias de medias estandarizadas se pueden reformular como logaritmos de los odds ratios al multiplicar por $\pi/\sqrt{3} = 1{,}814$. Una vez que las diferencias de medias estandarizadas (o los logaritmos de los odds ratios) y sus errores estándar se han calculado para todos los estudios en el metanálisis, es posible combinarlos mediante el método genérico del inverso de la varianza en RevMan. Los errores estándar se pueden calcular para todos los estudios entrando los datos en RevMan como tipo de datos de resultado dicotómicos o continuos, según sea apropiado, y convirtiendo los intervalos de confianza para los logaritmos de los odds ratios y las diferencias de medias estandarizadas en errores estándar (ver Capítulo 7, Sección 7.7.7.2).

9.4.7 Metanálisis de resultados ordinales y escalas de medición

Comúnmente los resultados ordinales y de escalas de medición se metanalizan como datos dicotómicos (de ser así, ver Sección 9.4.4) o como datos continuos (de ser así, ver Sección 9.4.5), según la forma en la cual los autores de los estudios realizaron los análisis originales.

En ocasiones es posible analizar los datos mediante los modelos de odds proporcionales cuando las escalas ordinales tienen un escaso número de categorías, se pueden obtener los números que se incluyen en cada categoría para cada grupo de intervención y se ha utilizado la misma escala ordinal en todos los estudios. Este enfoque puede hacer que el uso de todos los datos disponibles sea más eficiente que la dicotomización, pero requiere acceso a programas estadísticos y da lugar a estadísticos resumen para los cuales es difícil encontrar una significación clínica.

El modelo de odds proporcionales utiliza el odds ratio proporcional como medida del efecto de la intervención (Agresti 1996). Suponga que hay tres categorías que están ordenadas de manera que 1 es lo mejor y 3 es lo peor. Los datos se pueden dicotomizar de dos maneras. Es decir, la categoría 1 constituye el éxito y las categorías 2 y 3 el fracaso, o las categorías 1 y 2 constituyen el éxito y la 3 el fracaso. Un modelo de odds proporcionales supone que existe un odds ratio similar para ambas dicotomías de los datos. Por lo tanto, el odds ratio calculado del modelo de odds proporcionales se puede interpretar como el odds de éxito en la intervención experimental con respecto al control, independientemente de cómo se dividen las categorías ordenadas en éxito y fracaso. Hay métodos (específicamente los modelos de regresión logística policotómicos) disponibles en programas estadísticos avanzados para calcular las estimaciones del logaritmo del odds ratio y su error estándar en cada estudio y para realizar un metanálisis (Whitehead 1994).

Las estimaciones de los logaritmos de los odds ratios y sus errores estándar a partir de un modelo de odds proporcionales puede ser metanalizado mediante el método genérico del inverso de la varianza en RevMan (ver Sección 9.4.3.2). Pueden aplicarse los métodos de análisis de efectos fijos y aleatorios. Si la misma escala ordinal se utilizó en todos los estudios, pero en algunos informes se presentó como un resultado dicotómico, aún es posible incluir todos los estudios en el metanálisis. En el contexto del modelo de las tres categorías pudiera significar que para algunos estudios la categoría 1 constituye el éxito, mientras que para otros, ambas categorías, la 1 y la 2, constituyen un éxito. Hay métodos disponibles para tratar esta situación, así como para combinar los datos de las escalas que están relacionadas pero tienen diferentes definiciones para sus categorías (Whitehead 1994).

9.4.8 Metanálisis de recuentos y tasas

Cuando cada participante puede presentar un suceso y puede presentarlo más de una vez los resultados se pueden expresar como datos de recuento (ver Sección 9.2.5). Por ejemplo, "el número de accidentes cerebrovasculares" o "el número de visitas al hospital" son recuentos. Es posible que estos sucesos no hayan ocurrido en absoluto, pero si ocurren no hay teóricamente un número máximo de ocurrencias para un individuo.

Como se describe en el Capítulo 7 (Sección 7.7.5), los datos de recuento se pueden analizar mediante métodos para datos dicotómicos (ver Sección 9.4.4), continuos (ver Sección 9.4.5) y de tiempo hasta el suceso (ver Sección 9.4.9), y también se pueden analizar como datos de tasa.

Los datos de tasa aparecen si se miden los recuentos para cada participante junto con el tiempo durante el cual se observaron. Lo anterior es particularmente apropiado cuando los sucesos de recuento son poco frecuentes. Por ejemplo, una mujer puede presentar dos accidentes cerebrovasculares durante un período de seguimiento de dos años. Su tasa de accidente cerebrovascular es uno por año de seguimiento (o, de forma equivalente, 0,083 por mes de seguimiento). Las tasas se resumen de forma convencional a nivel grupal. Por ejemplo, los participantes de un grupo control de un ensayo aleatorio pueden presentar 85 accidentes cerebrovasculares durante un total de 2836 persona-años de seguimiento. Una suposición subyacente asociada con el uso de las tasas es que el riesgo de un suceso es constante entre los participantes y a lo largo del tiempo. Esta suposición se debe valorar cuidadosamente para cada situación. Por ejemplo, en los estudios de anticoncepción se han utilizado las tasas (conocidas como índices de Pearl) para describir el número de embarazos por 100 mujeres-años de seguimiento. Actualmente esto se considera inapropiado porque las parejas tienen riesgos diferentes de concepción, y el riesgo para cada mujer cambia con el tiempo. En estos momentos los embarazos se analizan con más frecuencia mediante los métodos de tablas de vida o de tiempo hasta el suceso, que investigan el tiempo que transcurre antes del primer embarazo.

Analizar los datos de recuentos como tasas no siempre es el enfoque más apropiado y es infrecuente en la práctica. Lo anterior se debe a:
1. es posible que la suposición de que existe un riesgo subyacente constante no sea apropiada; y
2. los métodos estadísticos no están tan desarrollados como para otros tipos de datos.

Los resultados de un estudio se pueden expresar como un cociente de tasas, que es el cociente de la tasa en el grupo de intervención experimental con respecto al cociente de la tasa en el grupo control. Suponga que ocurrieron E_E sucesos durante T_E participante-años de seguimiento en el grupo de intervención experimental, y E_C sucesos durante T_C participante-años de seguimiento en el grupo de intervención control. El cociente de tasas es

$$\text{cociente de tasas} = \frac{E_E/T_E}{E_C/T_C} = \frac{E_E T_C}{E_C T_E}.$$

Los logaritmos (naturales) de los cocientes de tasas se pueden combinar entre los estudios mediante el método genérico del inverso de la varianza (ver Sección 9.4.3.2). Un error estándar aproximado del logaritmo del cociente de tasas está dado por

$$\text{EE de ln cociente de tasas} = \sqrt{\frac{1}{E_E} + \frac{1}{E_C}}.$$

Se puede agregar una corrección de 0,5 a cada recuento en el caso de cero sucesos. Es de señalar que la selección de la unidad de tiempo (es decir, pacientes-meses, mujeres-años,

etc.) no es relevante porque se cancela en el cociente de tasas y no figura en el error estándar. Sin embargo, las unidades se deben mostrar cuando se presenten los resultados del estudio. Un medio alternativo para la estimación del cociente de tasas es a través del enfoque de Whitehead y Whitehead (Whitehead 1991).

En un ensayo aleatorizado, los cocientes de tasas a menudo pueden ser muy similares a los riesgos relativos obtenidos después de dicotomizar a los participantes, ya que el período promedio de seguimiento debe ser similar en todos los grupos de intervención. Sin embargo, los cocientes de tasas y los riesgos relativos serán diferentes si una intervención afecta la probabilidad de algunos participantes de presentar múltiples sucesos.

También es posible centrar la atención en la diferencia de tasas,

$$\text{diferencia de tasas} = \frac{E_E}{T_E} - \frac{E_C}{T_C}.$$

Un error estándar aproximado para la diferencia de tasas es

$$\text{EE de diferencia de tasas} = \sqrt{\frac{E_E}{T_E^{\,2}} + \frac{E_C}{T_C^{\,2}}}\,.$$

Nuevamente, el análisis requiere el uso del método genérico del inverso de la varianza en RevMan. Una de las pocas discusiones sobre el metanálisis de tasas, que aún es demasiado corta, es la realizada por Hasselblad y McCrory (Hasselblad 1995).

9.4.9 Metanálisis de resultados de tiempo hasta el suceso

Existen dos enfoques disponibles en RevMan para el metanálisis de tiempo hasta el suceso. El que se utilice dependerá de qué datos se extrajeron de los estudios primarios, o se obtuvieron del reanálisis de los datos de pacientes individuales.

Si se han obtenido los estadísticos “O-E” y “V”, a través del reanálisis de los datos de pacientes individuales o de los estadísticos agregados presentados en los informes del estudio, entonces estos estadísticos pueden ser entradas directamente en RevMan mediante el tipo de resultado “O-E y Varianza”. Existen varias formas de calcular los estadísticos “O-E” y “V”. El método de Peto aplicado a los datos dicotómicos (Sección 9.4.4.2) produce un odds ratio; un enfoque *log-rank* produce un cociente de riesgos instantáneo y una variación del método de Peto para el análisis de los datos de tiempo hasta el suceso produce algo entre ambos. La medida de efecto apropiada se debe especificar en RevMan. En RevMan solamente están disponibles métodos de metanálisis de efectos fijos para los resultados “O-E y V”.

Alternativamente, si se han obtenido las estimaciones de los logaritmos de los cocientes de riesgo instantáneos y los errores estándar a partir de modelos de regresión de riesgos proporcionales de Cox, los resultados del estudio se pueden combinar mediante el método genérico del inverso de la varianza (ver Sección 9.4.3.2). Están disponibles los análisis de efectos fijos y efectos aleatorios.

Si se obtiene una mezcla de las estimaciones log-rank y del modelo de Cox de los estudios, todos los resultados se pueden combinar mediante el método genérico del inverso de la varianza, ya que las estimaciones log-rank se pueden convertir en el logaritmo de los cocientes de riesgos instantáneos y errores estándar mediante la fórmula proporcionada en el Capítulo 7 (Sección 7.7.6).

9.4.10 Un resumen de los métodos de metanálisis disponibles en RevMan

La Tabla 9.4.a enumera las opciones para el análisis estadístico que están disponibles en RevMan. RevMan requiere que el revisor seleccione un método de preferencia para cada resultado. Si el mismo no se especifica, entonces el programa informático por defecto utiliza el odds ratio de Mantel-Haenszel de efectos fijos para los resultados dicotómicos, la diferencia de medias de efectos fijos para los resultados continuos y el modelo de efectos fijos para los resultados genéricos del inverso de la varianza. Es importante que los revisores aclaren qué método utilizan cuando se presentan los resultados en el texto de una revisión, ya que no es posible garantizar que un metanálisis mostrado al usuario coincida con el método preferido seleccionado.

Tabla 9.4.a: Resumen de los métodos de metanálisis disponibles en RevMan

Tipo de dato	Medida de efecto	Métodos de efectos fijos	Métodos de efectos aleatorios
Dicotómicos	Odds ratio (OR)	Mantel-Haenszel (M-H) Inverso de la Varianza (IV) Peto	Mantel-Haenszel (M-H) Inverso de la Varianza (IV)
	Razón de riesgos (RR)	Mantel-Haenszel (M-H) Inverso de la Varianza (IV)	Mantel-Haenszel (M-H) Inverso de la Varianza (IV)
	Diferencia de riesgos (DR)	Mantel-Haenszel (M-H) Inverso de la Varianza (IV)	Mantel-Haenszel (M-H) Inverso de la Varianza (IV)
Continuos	Diferencia de medias (DM)	Inverso de la Varianza (IV)	Inverso de la Varianza (IV)
	Diferencia de medias estandarizada (DME)	Inverso de la Varianza (IV)	Inverso de la Varianza (IV)
O-E y Varianza	*Especificado por el usuario* (por defecto "odds ratio de Peto")	Peto	*Ninguno*
Varianza inversa genérica	*Especificado por el usuario*	Inverso de la Varianza (IV)	Inverso de la Varianza (IV)
Otros datos	*Especificado por el usuario*	*Ninguno*	*Ninguno*

9.4.11 Uso del recuento de votos para el metanálisis

En ocasiones los metanálisis utilizan el "recuento de votos" para comparar el número de estudios positivos con el número de estudios negativos. El recuento de votos está limitado a responder la pregunta "¿hay alguna prueba de efecto?". Pueden aparecer dos problemas con el recuento de votos, lo que indica que se debería evitar siempre que fuera posible. Primero, aparecen problemas si se utilizan decisiones subjetivas o la significación estadística para

definir los estudios “positivos” y “negativos” (Cooper 1980, Antman 1992). Para realizar el recuento de votos correctamente el número de estudios que muestren daño se debe comparar con el número que muestre beneficio, independientemente de la significación estadística o el tamaño de sus resultados. Se puede utilizar un test de signos para evaluar la significación de las pruebas de existencia de un efecto en cualquier dirección (si no hay efecto los estudios se distribuirán de forma equilibrada alrededor de la hipótesis nula de ninguna diferencia). Segundo, el recuento de votos no toma en cuenta las ponderaciones diferenciales dadas a cada estudio. El recuento de votos se podria considerar un último recurso en situaciones donde no es posible aplicar los métodos metanalíticos estándar (como cuando no hay medidas de resultado consistentes).

9.5 Heterogeneidad

9.5.1 ¿Qué es heterogeneidad?

Es inevitable que los estudios que se agrupan en una revisión sistemática difieran. A cualquier tipo de variabilidad entre los estudios de una revisión sistemática se le puede llamar heterogeneidad. Puede ser útil distinguir los diferentes tipos de heterogeneidad. La variabilidad en los participantes, las intervenciones y los resultados estudiados se puede describir como diversidad clínica (algunas veces llamada heterogeneidad clínica), y la variabilidad en el diseño del estudio y el riesgo de sesgo se puede describir como diversidad metodológica (algunas veces llamada heterogeneidad metodológica). La variabilidad en los efectos de la intervención que se evalúan en diferentes estudios se conoce como heterogeneidad estadística y es consecuencia de la diversidad clínica o metodológica, o ambas, entre los estudios. La heterogeneidad estadística se manifiesta en que los efectos de la intervención observados son más diferentes entre sí que lo que sería de esperar si se debieran sólo al error aleatorio (azar). Aquí se seguirá la convención y a la heterogeneidad estadística se le llamará simplemente heterogeneidad.

La variabilidad clínica dará lugar a heterogeneidad si el efecto de la intervención está afectado por factores que varían entre los estudios; principalmente las características específicas de las intervenciones o los pacientes. En otras palabras, el verdadero efecto de la intervención será diferente en estudios diferentes.

Es de esperar que las diferencias entre los estudios en cuanto a factores metodológicos como el uso del cegamiento y la ocultación de la asignación, o si existen diferencias entre los estudios en cuanto a la forma en que se definen y miden los resultados, den lugar a diferencias en los efectos observados de la intervención. La heterogeneidad estadística significativa que surge de la diversidad metodológica o las diferencias en las evaluaciones de resultados indica que no todos los estudios están calculando la misma cantidad, pero no indica necesariamente que el efecto verdadero de la intervención varíe. En particular, la heterogeneidad asociada solamente con la diversidad metodológica indicaría que los estudios presentan diferentes grados de sesgo. Las pruebas empíricas indican que algunos aspectos del diseño pueden afectar el resultado de los ensayos clínicos, aunque éste no siempre es el caso. Se realiza una discusión más amplia al respecto en el Capítulo 8.

El alcance de una revisión determinará en gran medida el grado en el cual los estudios incluidos en una revisión son diferentes. Algunas veces una revisión incluirá estudios que abordan varias preguntas, por ejemplo, cuando hay varias intervenciones diferentes de interés para la misma afección (ver también Capítulo 5, Sección 5.6). Los estudios de cada intervención se deben analizar y presentar por separado. El metanálisis sólo se debe considerar cuando un grupo de estudios es suficientemente homogéneo en cuanto a los participantes, las intervenciones y los resultados, para proporcionar un resumen significativo. A menudo es apropiado tener una perspectiva más amplia en un metanálisis que en un ensayo clínico único. Una analogía frecuente es que las revisiones sistemáticas juntan manzanas y naranjas, y que la combinación de las mismas puede dar a lugar a un resultado

sin sentido. Lo anterior es cierto si hay un interés intrínseco en las manzanas y naranjas por sí mismas, pero pudiera no serlo si se utilizan para responder a una pregunta más amplia acerca de las frutas. Por ejemplo, un metanálisis puede evaluar razonablemente el efecto promedio de una clase de fármaco al combinar los resultados de ensayos donde cada uno evalúe el efecto de un fármaco diferente de la clase.

Una revisión puede tener el interés específico de investigar cómo los aspectos clínicos y metodológicos de un estudio se relacionan con sus resultados. De ser posible estas investigaciones se deben especificar *a priori*, es decir, en el protocolo de la revisión sistemática. Es correcto que una revisión sistemática se centre en el examen de la relación entre alguna/s característica/s de los estudios y el tamaño del efecto de la intervención, en lugar de en obtener una estimación resumen del efecto entre varios estudios (ver Sección 9.6). Puede ser mejor utilizar la metarregresión para estos fines, aunque no está implementada en RevMan (ver Sección 9.6.4).

9.5.2 Identificación y medición de la heterogeneidad

Es importante considerar hasta qué grado los resultados de los estudios son consistentes. Si los intervalos de confianza para los resultados de los estudios individuales (generalmente mostrados gráficamente mediante líneas horizontales) tienen escasa superposición, habitualmente indica la presencia de heterogeneidad estadística. De manera más formal, existe un test estadístico de heterogeneidad. La prueba de ji cuadrado (χ^2, o Chi^2) se incluye en los diagramas de árbol de las revisiones Cochrane. La misma evalúa si las diferencias observadas en los resultados son compatibles con el azar. Un valor de p bajo (o un estadístico ji cuadrado grande con respecto a sus grados de libertad) proporciona indicios de heterogeneidad en los efectos de la intervención (variación en las estimaciones más allá del azar).

Se debe tener precaución al interpretar la prueba de ji cuadrado, ya que la misma tiene una potencia estadística baja en la situación (frecuente) de un metanálisis en el cual los estudios tienen un tamaño de muestra pequeño o son pocos en número. Lo anterior significa que aunque un resultado estadísticamente significativo puede indicar un problema con la heterogeneidad, un resultado no significativo no se debe tomar como prueba de falta de heterogeneidad. Es también por este motivo que un valor de p de 0,10, en lugar del 0,05 convencional, se utiliza algunas veces para determinar la significación estadística. Un problema adicional con esta prueba, que ocurre a menudo en las revisiones Cochrane, es que cuando hay muchos estudios en un metanálisis, la prueba tiene una potencia estadística grande para detectar una pequeña cantidad de heterogeneidad que pudiera no ser clínicamente importante.

Algunos argumentan que, como siempre existe diversidad clínica y metodológica en un metanálisis, la heterogeneidad estadística es inevitable (Higgins 2003). Por lo tanto, el test de heterogeneidad no es relevante para la selección del análisis; la heterogeneidad siempre existirá, se tenga o no la capacidad de detectarla mediante un test estadístico. Se han desarrollado métodos para cuantificar la inconsistencia entre los estudios, lo que hace que cambie el foco de probar si hay heterogeneidad hacia evaluar su impacto en el metanálisis. Un estadístico útil para cuantificar la inconsistencia es

$$I^2 = \left(\frac{Q - df}{Q} \right) \times 100\%,$$

donde Q es el estadístico ji cuadrado y gl son los grados de libertad (Higgins 2002, Higgins 2003). El estadistico I^2 describe el porcentaje de la variabilidad en las estimaciones del efecto que se debe a la heterogeneidad en lugar de al error de muestreo (azar).

Establecer umbrales para la interpretación de I^2 puede confundir, ya que la importancia de la inconsistencia depende de varios factores. Una guía aproximada para su interpretación es la siguiente:

- 0% al 40%: pudiera no ser importante;
- 30% al 60%: puede representar heterogeneidad moderada *;
- 50% a 90%: puede representar heterogeneidad significativa*;
- 75% al 100%: heterogeneidad considerable*.

*La importancia del valor de I^2 observado depende de (i) la magnitud y dirección de los efectos y (ii) la fuerza de los tests de heterogeneidad (p.ej. el valor de p de la prueba de ji cuadrado, o un intervalo de confianza para I^2).

9.5.3 Estrategias para abordar la heterogeneidad

Hay varias opciones disponibles si se identifica heterogeneidad (estadística) entre un grupo de estudios que de otra manera se considerarían adecuados para un metanálisis.

1. Verifique nuevamente si los datos son correctos
La heterogeneidad considerable puede indicar que los datos se han extraído o entrado en RevMan de forma incorrecta. Por ejemplo, si los errores estándar se grabaron erróneamente como desviaciones estándar para los resultados continuos, se pudiera manifestar como intervalos de confianza exageradamente estrechos con una superposición pequeña y, por lo tanto, heterogeneidad significativa. Los errores de la unidad de análisis también pueden ser causa de heterogeneidad (ver Sección 9.3).

2. No realice un metanálisis
No es necesario que una revisión sistemática contenga metanálisis (O'Rourke 1989). Si existe una variabilidad considerable en los resultados y, particularmente, si existe inconsistencia en la dirección del efecto, puede ser erróneo citar un valor promedio para el efecto de la intervención.

3. Explore la heterogeneidad
No hay dudas de que es muy importante determinar las causas de la heterogeneidad entre los resultados de los estudios. Este proceso es problemático porque a menudo hay muchas características que varían entre los estudios y que se pueden seleccionar. La heterogeneidad se puede explorar mediante la realización de análisis de subgrupos (ver Sección 9.6.3) o una metarregresión (ver Sección 9.6.4), aunque este último método no está implementado en RevMan. Idealmente las investigaciones de las características de los estudios que se pueden asociar con heterogeneidad se deben preespecificar en el protocolo de una revisión (ver Sección 9.1.7). Sólo es posible establecer conclusiones confiables de los análisis que fueron verdaderamente preespecificados antes de inspeccionar los resultados de los estudios, e incluso estas conclusiones se deben interpretar con precaución. En la práctica los revisores a menudo estarán familiarizados con algunos resultados de los estudios cuando redacten el protocolo, por lo que no es posible una verdadera preespecificación. Las exploraciones de la heterogeneidad que se recomiendan después de que se ha identificado la existencia de heterogeneidad sólo pueden, en el mejor de los casos, llevar a la generación de hipótesis. Se deben interpretar incluso con más precaución y en general no se deben enumerar entre las conclusiones de una revisión. Además, las exploraciones de la heterogeneidad cuando hay muy pocos estudios tienen un valor cuestionable.

4. Ignore la heterogeneidad
Los metanálisis de efectos fijos ignoran la heterogeneidad. La estimación combinada del efecto (*pooled estimate*) de un metanálisis de efectos fijos se interpreta generalmente como la mejor estimación del efecto de la intervención. Sin embargo, la presencia de heterogeneidad indica que es posible que no exista un único efecto de la intervención, sino una distribución de efectos de la intervención. Por lo tanto, la estimación agrupada de los efectos fijos puede ser un efecto de la intervención que no existe en realidad en una población y, por lo tanto, tiene un intervalo de confianza que no tiene sentido y que además

es muy estrecho (ver Sección 9.5.4). Sin embargo, el valor de p obtenido de un metanálisis de efectos fijos proporciona un test útil de la hipótesis nula de que no hay efecto en los estudios.

5. Realice un metanálisis de efectos aleatorios
Un metanálisis de efectos aleatorios se puede utilizar para incorporar la heterogeneidad entre los estudios. El mismo no es un sustituto de una exploración cuidadosa de la heterogeneidad, cuyo objetivo principal es la heterogeneidad que no se puede explicar. En la Sección 9.5.4 aparece una discusión ampliada de esta opción.

6. Cambie la medida de efecto
La heterogeneidad puede ser una consecuencia artificial de la selección inapropiada de la medida de efecto. Por ejemplo, cuando los estudios obtienen datos de resultado continuos mediante escalas o unidades diferentes puede haber una heterogeneidad extrema evidente cuando se utiliza la diferencia de medias pero no cuando se utiliza la diferencia de medias estandarizada, más apropiada. Además, la selección de la medida de efecto para los resultados dicotómicos (odds ratio, razón de riesgos o diferencia de riesgos) puede afectar el grado de heterogeneidad entre los resultados. En particular, cuando los riesgos del grupo control varían, los odds ratios o las razones de riesgos homogéneos necesariamente darán lugar a diferencias de riesgos heterogéneas, y *viceversa*. Sin embargo, aún no está claro si la homogeneidad del efecto de la intervención en un metanálisis particular es un criterio apropiado para la selección entre estas medidas (ver también Sección 9.4.4.4).

7. Excluya estudios
La heterogeneidad se puede deber a la presencia de uno o dos estudios atípicos con resultados contradictorios con respecto a los estudios restantes. En general no es recomendable excluir estudios de un metanálisis sobre la base de sus resultados, ya que puede introducir sesgo. Sin embargo, si hay un motivo evidente para el resultado atípico, el estudio se pudiera eliminar con más confianza. Como habitualmente es posible encontrar al menos una característica de algún estudio en un metanálisis que lo hace diferente de otros, este criterio no es fiable porque es demasiado fácil de cumplir. Es aconsejable realizar los análisis con y sin los estudios atípicos como parte de un análisis de sensibilidad (ver Sección 9.7). Cuando sea posible, se deben especificar en el protocolo las posibles fuentes de diversidad clínica que pueden dar lugar a estas situaciones.

9.5.4 Incorporación de la heterogeneidad en los modelos de efectos aleatorios

Un metanálisis de efectos fijos proporciona un resultado que se puede considerar un “efecto típico de la intervención” de los estudios incluidos en el análisis. Para calcular un intervalo de confianza para un metanálisis de efectos fijos se supone que el efecto verdadero de la intervención (en cuanto a la magnitud y la dirección) es el mismo valor en cada estudio (es decir, es fijo entre los estudios). Esta suposición implica que las diferencias observadas entre los resultados de los estudios se deben solamente al azar, es decir que no hay heterogeneidad estadística.

Cuando existe heterogeneidad que no es posible explicar fácilmente, un enfoque analítico es incorporarla en un modelo de efectos aleatorios. Un modelo de metanálisis de efectos aleatorios incluye la suposición de que los efectos que se calculan en los diferentes estudios no son idénticos, pero siguen alguna distribución. El modelo representa la falta de conocimiento acerca de por qué los efectos reales o aparentes de la intervención difieren al considerar las diferencias como si las mismas fueran aleatorias. El centro de esta distribución describe el promedio de los efectos, mientras que su amplitud describe el grado de heterogeneidad. La selección convencional de la distribución es una distribución normal. Es difícil establecer la validez de cualquier suposición relacionada con la distribución, y es una crítica frecuente a los metanálisis de efectos aleatorios. No se conoce la importancia de la forma particular supuesta para esta distribución.

Es de señalar que un modelo de efectos aleatorios no "toma en cuenta" la heterogeneidad, en el sentido de que deja de ser un problema. Siempre es aconsejable explorar las posibles causas de heterogeneidad, aunque es posible que no se disponga de suficientes estudios como para hacerlo de forma adecuada (ver Sección 9.6).

En los análisis de efectos aleatorios en RevMan, la estimación combinada y el intervalo de confianza se refieren al centro de la distribución de los efectos de la intervención, pero no describen la amplitud de la distribución. A menudo la estimación agrupada y su intervalo de confianza se citan de manera aislada como una estimación alternativa de la magnitud evaluada en un metanálisis de efectos fijos, lo cual no es apropiado. El intervalo de confianza de un metanálisis de efectos aleatorios describe la incertidumbre en la ubicación de la media de efectos sistemáticamente diferentes en estudios diferentes. No describe el grado de heterogeneidad entre los estudios, como se cree frecuentemente. Por ejemplo, cuando se incluyen muchos estudios en un metanálisis, es posible obtener un intervalo de confianza estrecho alrededor de la estimación de efectos aleatorios del efecto medio, incluso cuando existe una gran cantidad de heterogeneidad.

Al igual que otros programas informáticos de metanálisis, RevMan presenta una estimación de la varianza entre estudios en un metanálisis de efectos aleatorios (conocido como tau cuadrado [τ^2 o Tau^2]). La raíz cuadrada de este número (es decir, tau) es la estimación de la desviación estándar de los efectos subyacentes entre los estudios. Para las medidas absolutas de efecto (p.ej. la diferencia de riesgos, la diferencia de medias, la diferencia de medias estandarizada), se puede obtener un rango aproximado del 95% de los efectos subyacentes al crear un intervalo a partir de 2xtau por debajo de la estimación combinada de los efectos aleatorios, hasta 2xtau por encima de la misma. Para las medidas relativas (p.ej. odds ratio, razón de riesgos), el intervalo se debe centrar en el logaritmo natural de la estimación combinada y se debe calcular el antilogaritmo de los límites (exponenciados) para obtener un intervalo en la escala del cociente. Se han propuesto intervalos alternativos para el efecto predicho en un estudio nuevo (Higgins 2008a). Es posible considerar que el rango de los efectos de la intervención observados en los estudios proporciona una idea general de la amplitud de la distribución de los efectos verdaderos de la intervención, pero de hecho la misma será ligeramente más amplia, ya que también describe el error aleatorio en las estimaciones del efecto observadas.

Si se considera que la variabilidad de los efectos (heterogeneidad estadística) se debe a la diversidad clínica, la estimación combinada de efectos aleatorios se debe interpretar de forma diferente de la estimación de efectos fijos porque se relaciona con un aspecto diferente. La estimación de efectos aleatorios y su intervalo de confianza abordan el problema "¿cuál es el efecto promedio de la intervención?", mientras que la estimación de efectos fijos y su intervalo de confianza abordan el problema "¿cuál es la mejor estimación del efecto de la intervención?". Las respuestas a estos interrogantes coinciden cuando no hay heterogeneidad o cuando la distribución de los efectos de la intervención es en general simétrica. Cuando las respuestas no coinciden es posible que la estimación de efectos aleatorios no refleje el efecto real en ninguna de las poblaciones específicas estudiadas.

La diversidad metodológica da lugar a heterogeneidad a través de los sesgos que afectan de forma variable los resultados de diferentes estudios. La estimación combinada de efectos aleatorios sólo calculará el efecto promedio del tratamiento si los sesgos están simétricamente distribuidos y causan una mezcla de sobrestimaciones y subestimaciones del efecto, lo que es poco probable que ocurra. En la práctica puede ser muy difícil diferenciar si la heterogeneidad está provocada por la diversidad clínica o la metodológica y en la mayoría de los casos es probable que se deba a ambas, por lo que es difícil hacer estas diferenciaciones en la interpretación.

Para cualquier conjunto específico de estudios en los cuales exista heterogeneidad, un intervalo de confianza alrededor de la estimación combinada de efectos aleatorios es más amplio que un intervalo de confianza alrededor de la estimación de efectos fijos. Lo anterior

sucederá si el estadístico I^2 es mayor que cero, incluso si no se detecta heterogeneidad con el test de ji cuadrado de heterogeneidad (Higgins 2003) (ver Sección 9.5.2). La decisión entre un metanálisis de efectos fijos y uno de efectos aleatorios nunca se debe hacer sobre la base de un test estadístico de heterogeneidad.

En un conjunto heterogéneo de estudios, un metanálisis de efectos aleatorios dará relativamente más ponderación a los estudios más pequeños que la que recibirían en un metanálisis de efectos fijos. Lo anterior se debe a que los estudios pequeños son más informativos sobre la distribución de los efectos entre los estudios que sobre un supuesto efecto común de la intervención. Se debe tener en cuenta que los análisis de efectos aleatorios sólo se aplican cuando se puede justificar la idea de una distribución "aleatoria" de los efectos de la intervención. En particular, si los resultados de los estudios más pequeños son sistemáticamente diferentes de los resultados de los estudios más grandes, lo cual puede suceder como resultado del sesgo de publicación o del sesgo dentro de los estudios más pequeños ((Egger 1997, Poole 1999, Kjaergard 2001), entonces un metanálisis de efectos aleatorios incrementará los efectos del sesgo (ver también Capítulo 10, Sección 10.4.4.1). Un metanálisis de efectos fijos tendrá una afectación menor, aunque no sería apropiado. En esta situación se recomienda no presentar metanálisis alguno o realizar un análisis de sensibilidad en el cual se excluyan los estudios más pequeños.

De manera similar, cuando hay poca información debido a que hay escasos estudios o a que los estudios son pequeños y tienen pocos sucesos, un análisis de efectos aleatorios proporcionará estimaciones deficientes de la amplitud de la distribución de los efectos de la intervención. El método de Mantel-Haenszel proporcionará estimaciones más sólidas del efecto promedio de la intervención, pero al costo de ignorar la heterogeneidad observada.

RevMan implementa una versión del metanálisis de efectos aleatorios descrita por DerSimonian y Laird (DerSimonian 1986). El atractivo de este método es que los cálculos son sencillos, pero tiene la desventaja teórica de que los intervalos de confianza son ligeramente demasiado estrechos como para incluir toda la incertidumbre derivada de haber estimado el grado de heterogeneidad. Existen métodos alternativos que incluyen toda la incertidumbre pero requieren de programas informáticos más avanzados (ver también Capítulo 16, Sección 16.8). En la práctica es probable que la diferencia en los resultados sea pequeña, a menos que haya pocos estudios. Para datos dicotómicos RevMan implementa dos versiones del modelo de efectos aleatorios de DerSimonian y Laird (ver Sección 9.4.4.3).

9.6 Investigación de la heterogeneidad

9.6.1 Interacción y modificación del efecto

¿El efecto de la intervención varía con poblaciones diferentes o con las características de la intervención (como dosis y duración)?. Los estadísticos llaman a esta variación interacción y los epidemiólogos la llaman modificación del efecto. Los métodos para investigar dichas interacciones incluyen los análisis de subgrupos y la metarregresión. Todos los métodos tienen limitaciones importantes.

9.6.2 ¿Qué son los análisis de subgrupos?

Los análisis de subgrupos consisten en separar a todos los participantes en subgrupos, a menudo para hacer comparaciones entre ellos. Los análisis de subgrupos se pueden realizar para subconjuntos de participantes (como hombres y mujeres), o para subconjuntos de estudios (como lugares geográficos diferentes). Los análisis de subgrupos se pueden realizar como una forma de investigar resultados heterogéneos o responder preguntas específicas acerca de grupos de pacientes, tipos de intervención o tipos de estudios particulares.

Los análisis de subgrupos de subconjuntos de participantes dentro de los estudios no son frecuentes en las revisiones sistemáticas de la bibliografía porque pocas veces se publican en los informes detalles suficientes para extraer por separado datos de los subconjuntos de participantes. Por el contrario, estos subconjuntos de participantes se analizan con facilidad cuando se han obtenido los datos de pacientes individuales (ver Capítulo 18). Los métodos que se describen en la Sección 9.6.3 son para subgrupos de ensayos.

Los hallazgos de múltiples análisis de subgrupos pueden dar lugar a confusión. Los análisis de subgrupos son observacionales por naturaleza y no se basan en comparaciones aleatorias. La probabilidad de falsos positivos y falsos negativos en los tests de significación aumenta rápidamente cuantos más análisis de subgrupos se realizan. Si sus hallazgos se presentan como conclusiones definitivas, existe un riesgo claro de que a los pacientes se les niegue una intervención efectiva o reciban tratamiento con una intervención inefectiva (o incluso perjudicial). Los análisis de subgrupos también pueden generar recomendaciones erróneas acerca de las lineas de investigaciones futuras que, si se siguen, pudieran malgastar recursos escasos.

Es de utilidad diferenciar entre los conceptos “interacción cualitativa” e “interacción cuantitativa” (Yusuf 1991). La interacción cualitativa existe si la dirección del efecto se invierte, es decir, si una intervención es beneficiosa en un subgrupo pero perjudicial en otro. La interacción cualitativa es poco frecuente. Se puede utilizar como argumento que el resultado más apropiado de un metanálisis es el efecto general entre todos los subgrupos. La interacción cuantitativa existe cuando varía el tamaño del efecto pero no su dirección, es decir, si una intervención es beneficiosa en un grado diferente en subgrupos distintos.

Los revisores encontrarán recomendaciones útiles con respecto a los análisis de subgrupos en Oxman y Guyatt (Oxman 1992) y Yusuf y cols. (Yusuf 1991). Ver también Sección 9.6.6.

9.6.3 Realización de los análisis de subgrupos

Los análisis de subgrupos se pueden realizar con RevMan. Se permiten los metanálisis dentro de los subgrupos y los metanálisis que combinan varios subgrupos. Es tentador comparar las estimaciones del efecto en diferentes subgrupos al considerar los resultados del metanálisis de cada subgrupo por separado. Lo anterior sólo se debe hacer de manera informal al comparar las magnitudes de los efectos. Observar que el efecto o el test de heterogeneidad en un subgrupo son estadísticamente significativos mientras que en el otro subgrupo no lo son, no indica que el factor subgrupo explique la heterogeneidad. Como es probable que diferentes subgrupos contengan una cantidad diferente de información y, por lo tanto, diferente capacidad para detectar los efectos, es extremadamente erróneo comparar solamente la significación estadística de los resultados.

9.6.3.1 ¿El efecto es diferente en subgrupos diferentes?

Una investigación válida de si una intervención funciona de forma diferente en subgrupos distintos incluye comparar los subgrupos entre sí. Cuando hay solamente dos subgrupos se puede considerar la superposición de los intervalos de confianza de las estimaciones resumen en los dos grupos. La falta de superposición de los intervalos de confianza indica significación estadística, pero es de señalar que los intervalos de confianza se pueden superponer en un grado pequeño y la diferencia ser aún estadísticamente significativa.

Un enfoque sencillo para un test de significación que se puede utilizar para investigar las diferencias entre dos o más subgrupos se describe en Deeks y cols. (Deeks 2001). Este método está implementado en RevMan para los análisis de efectos fijos según el método del inverso de la varianza. Si se utilizan los métodos de Mantel-Haenszel para el tipo de datos dicotómicos, entonces el test presentaría una ligera inexactitud debido a la forma en la cual se calculó el estadístico ji cuadrado para la heterogeneidad. El procedimiento se basa en el test del estadístico ji cuadrado de heterogeneidad que aparece en la esquina inferior izquierda

de los diagramas de árbol, y es la siguiente. Suponga que un estadístico ji cuadrado de heterogeneidad Q_{tot} está disponible para todos los estudios, y que los test estadísticos de heterogeneidad Q_1 hasta Q_J están disponibles para J subgrupos (de manera que cada estudio está en un subgrupo y sólo uno). Entonces, el nuevo estadístico $Q_{int} = Q_{tot} -(Q_1+ \dots +Q_J)$, comparado con una distribución de ji cuadrado con J-1 grados de libertad, testea las diferencias entre los subgrupos. Una alternativa más flexible para probar las diferencias entre los subgrupos es utilizar las técnicas de metarregresión, en las cuales se permite la heterogeneidad residual (es decir, la heterogeneidad no explicada por el subagrupamiento) (ver Sección 9.6.4). Este enfoque se puede considerar preferible debido al alto riesgo de resultados falsos positivos cuando se comparan subgrupos en un modelo de efectos fijos (Higgins 2004).

9.6.4 Metarregresión

La división de los estudios en subgrupos (ver Sección 9.6.2) se puede considerar como una investigación de cómo una característica categórica de un estudio se asocia con los efectos de la intervención en el metanálisis. Por ejemplo, los estudios donde la ocultación de la secuencia de asignación fue adecuada pueden producir resultados diferentes a los de los estudios que tuvieron una asignación inadecuada. Aquí, la ocultación de la secuencia de asignación, ya sea adecuada o inadecuada, es una característica categórica a nivel de estudio. La metarregresión es una extensión del análisis de subgrupos que permite investigar el efecto de características categóricas y continuas, y en principio permite investigar simultáneamente los efectos de múltiples factores (aunque es poco frecuente, debido al número insuficiente de estudios) (Thompson 2002). En general la metarregresión no se debe considerar cuando hay menos de diez estudios en un metanálisis.

Las metarregresiones son similares en esencia a las regresiones simples, en las cuales una variable de resultado es predicha según los valores de una o más variables explicativas. En la metarregresión la variable de resultado es la estimación del efecto (por ejemplo, una diferencia de medias, una diferencia de riesgos, un logaritmo del odds ratio o un logaritmo de la razón de riesgos). Las variables explicativas son características de los estudios que pudieran influir en el tamaño del efecto de la intervención. A menudo se les llama “modificadores potenciales del efecto” o covariables. Habitualmente las metarregresiones difieren de las regresiones simples en dos aspectos. Primero, los estudios más grandes tienen una mayor influencia con respecto a los estudios más pequeños, ya que los estudios se ponderan según la precisión de su estimación del efecto. Segundo, es recomendable permitir que la heterogeneidad residual entre los efectos de la intervención no se modele junto con las variables explicativas. Lo anterior da lugar al término “metarregresión de efectos aleatorios”, ya que la variabilidad extra se incorpora de la misma manera que en el metanálisis de efectos aleatorios (Thompson 1999).

El coeficiente de regresión obtenido de un análisis de metarregresión describirá cómo la variable de resultado (el efecto de la intervención) cambia con el aumento de una unidad en la variable explicativa (el modificador potencial del efecto). La significación estadística del coeficiente de regresión es un test de si existe una relación lineal entre el efecto de la intervención y la variable explicativa. Si el efecto de la intervención es una medida de cociente, siempre se debe utilizar el valor del efecto de la intervención transformado logarítmicamente en el modelo de regresión (ver Sección 9.2.7), y el exponencial del coeficiente de regresión dará una estimación del cambio relativo en el efecto de la intervención con el aumento de una unidad en la variable explicativa.

La metarregresión también se puede utilizar para investigar diferencias en variables categóricas explicativas como se hace en los análisis de subgrupos. Si hay J subgrupos, la pertenencia a un subgrupo particular se indica mediante J-1 variables simuladas (que solamente pueden tomar valores de cero o uno) en el modelo de metarregresión (como en el modelo de regresión lineal estándar). Los coeficientes de regresión estimarán cómo el efecto de la intervención en cada subgrupo difiere con respecto a un subgrupo tomado como

referencia. El valor de p de cada coeficiente de regresión indicará si esta diferencia es estadísticamente significativa.

La metarregresión se puede realizar mediante el macro “metareg” disponible en el programa informático Stata.

9.6.5 Selección de las características del estudio para los análisis de subgrupos y la metarregresión

Los revisores deben tener precaución al realizar análisis de subgrupos y al interpretar cualquier análisis que realicen. Aquí se esbozan algunas consideraciones para la selección de las características (también llamadas variables explicativas, modificadores potenciales del efecto o covariables) que se investigarán con respecto a su posible influencia sobre el tamaño del efecto de la intervención. Estas consideraciones se aplican de manera similar a los análisis de subgrupos y las metarregresiones. Se pueden obtener detalles adicionales de Oxman y Guyatt (Oxman 1992) y Berlin y Antman (Berlin 1994).

9.6.5.1 Asegúrese de que haya estudios adecuados para justificar los análisis de subgrupos y las metarregresiones

Es muy poco probable que una investigación de la heterogeneidad produzca hallazgos útiles, a menos que haya un número significativo de estudios. Vale la pena señalar el consejo típico para realizar análisis de regresión simples: que deben estar disponibles al menos diez observaciones (es decir, diez estudios o metanálisis) para cada característica a la que se le aplique el modelo. Sin embargo, incluso éstas pueden ser demasiado pocas cuando las covariables no están distribuidas de forma equilibrada.

9.6.5.2 Especifique de antemano las características

De ser posible, los revisores deben preespecificar en el protocolo las características que posteriormente se incluirán en los análisis de subgrupos o la metarregresión. Al preespecificar las características se reduce la probabilidad de hallazgos espurios, en primer lugar porque limita el número de subgrupos investigados y en segundo lugar porque evita que el conocimiento de los resultados de los estudios influya en qué subgrupos se analicen. La verdadera preespecificación es difícil en las revisiones sistemáticas porque a menudo se conocen los resultados de algunos estudios relevantes cuando se redacta el protocolo. Si una característica se pasó por alto en el protocolo, pero es claramente de gran importancia y se justifica por evidencia externa, entonces los revisores no deben rechazar la exploración de la misma. Sin embargo, estos análisis post hoc se deben identificar como tales.

9.6.5.3 Seleccione un pequeño número de características

La probabilidad de un resultado falso positivo entre los análisis de subgrupos y la metarregresión aumenta con el número de características investigadas. Es difícil indicar un número máximo de características a examinar, especialmente porque el número de estudios disponibles no se conoce por adelantado. Si se investigan más de una o dos características, pudiera ser recomendable ajustar el nivel de significación para tomar en cuenta las comparaciones múltiples. Se recomienda la ayuda de un estadístico (ver Capítulo 16, Sección 16.7)

9.6.5.4 Asegúrese de que hay una justificación científica para investigar cada característica

La selección de las características debe estar motivada por hipótesis biológicas y clínicas, idealmente apoyadas por pruebas provenientes de fuentes diferentes a los estudios incluidos. Es poco probable que los análisis de subgrupos que utilizan características que no son factibles ni clínicamente relevantes sean útiles y se deben evitar. Por ejemplo, una relación

entre el efecto de la intervención y el año de publicación pocas veces es en sí misma clínicamente informativa, y si es estadísticamente significativa corre el riesgo de iniciar una "pesca" de datos *post hoc* de factores que pueden haber cambiado con el tiempo.

Los factores pronósticos son los que predicen el resultado de una enfermedad o afección, mientras que los modificadores del efecto son los factores que influyen en cuán bien funciona una intervención para afectar un resultado. Es frecuente confundir los factores pronósticos y los modificadores del efecto cuando se planifican los análisis de subgrupos, especialmente en el estadio de protocolo. Los factores pronósticos no son buenos candidatos para los análisis de subgrupos, a menos que se considere que modifican el efecto de la intervención. Por ejemplo, ser fumador puede ser un fuerte predictor de mortalidad en los próximos diez años, pero es posible que no haya un motivo para que influya en el efecto de un tratamiento farmacológico sobre la mortalidad (Deeks 1998). Los modificadores potenciales del efecto pueden incluir las intervenciones precisas (dosis de tratamiento activo, selección del tratamiento de comparación), cómo se realizó el estudio (duración del seguimiento) o la metodología (diseño y calidad).

9.6.5.5 Tenga en cuenta que no siempre es posible identificar el efecto de una característica

Muchas características que podrían tener efectos importantes sobre cuán bien funciona una intervención no se pueden investigar mediante análisis de subgrupos o metarregresión. Se trata de las características de los participantes que podrían variar significativamente dentro de los estudios, pero que sólo se pueden resumir a nivel de estudio. Un ejemplo es la edad. Considere un grupo de ensayos clínicos que incluyan adultos entre 18 y 60 años de edad. Puede haber una fuerte relación entre la edad y el efecto de la intervención que es evidente dentro de cada estudio. Sin embargo, si la media de las edades en los ensayos es similar, entonces no habrá una relación evidente al examinar las medias de las edades de los ensayos y las estimaciones del efecto a nivel de ensayo. El problema surge al agregar los resultados de los individuos y se conoce de varias formas como sesgo de agregación, sesgo ecológico o falacia ecológica (Morgenstern 1982, Greenland 1987, Berlin 2002). Incluso es posible que las diferencias entre los estudios muestren el patrón opuesto al observado dentro de cada estudio.

9.6.5.6 Considere si la característica está muy relacionada con otra característica (factor de confusión)

El problema de la "confusión" complica la interpretación de los análisis de subgrupos y las metarregresiones y puede dar lugar a conclusiones incorrectas. Dos características son factores de confusión si no es posible separar sus influencias sobre el efecto de la intervención. Por ejemplo, si coincide que los estudios que implementan una versión intensiva de un tratamiento son los que reclutaron pacientes con la enfermedad más grave, entonces no es posible decir qué aspecto es la causa de cualquier diferencia en las estimaciones del efecto entre estos estudios y otros. En la metarregresión la colinearidad entre los modificadores potenciales del efecto provoca dificultades similares, como señalan Berlin y Antman (Berlin 1994). El cálculo de las correlaciones entre las características de los estudios proporcionará alguna información acerca de qué características del estudio puede ser un factor de confusión con respecto a otra.

9.6.6 Interpretación de los análisis de subgrupos y las metarregresiones

Es necesario tomar precauciones para interpretar de manera apropiada los análisis de subgrupos y las metarregresiones. Para una discusión más detallada ver Oxman y Guyatt (Oxman 1992).

- La comparaciones de subgrupos son observacionales

Se debe recordar que los análisis de subgrupos y las metarregresiones son de naturaleza completamente observacional. Estos análisis investigan las diferencias entre los estudios. Incluso si los individuos se asignan al azar a uno u otro grupo dentro de un ensayo clínico, no se asignaron al azar para participar en uno u otro ensayo. Por lo tanto, los análisis de subgrupos presentan las limitaciones de cualquier investigación observacional, incluido el posible sesgo debido a la confusión con otras características a nivel de estudio. Además, incluso una diferencia real entre los subgrupos no se debe necesariamente a la clasificación de los subgrupos. Como ejemplo, un análisis de subgrupos de trasplante de médula ósea para tratar la leucemia pudiera mostrar una fuerte asociación entre la edad del hermano donante y el éxito del trasplante. Sin embargo, es probable que este hallazgo no signifique que la edad del donante sea importante. De hecho, la edad de la persona receptora probablemente sea un factor clave y el hallazgo de subgrupo se deba sencillamente a la fuerte asociación entre la edad del receptor y la edad de su hermano.

- ¿El análisis fue preespecificado o post hoc?

Los revisores deben señalar si los análisis de subgrupos fueron preespecificados o se realizaron después de que se habían recopilado los resultados de los estudios (*post hoc*). Se debe tener más confianza en un análisis de subgrupos si el mismo fue uno de los pocos análisis preespecificados. La realización de varios análisis de subgrupos *post hoc* para explicar la heterogeneidad es “pescar” los datos. La “pesca” de datos está criticada porque habitualmente es posible encontrar una explicación aparente pero falsa al considerar muchas características diferentes.

- ¿Existen pruebas indirectas que apoyan los hallazgos?

Las diferencias entre los subgrupos deben ser clínicamente factibles y estar apoyadas por otras pruebas externas o indirectas para que sean convincentes.

- ¿La magnitud de la diferencia es importante en la práctica?

Si la magnitud de una diferencia entre subgrupos no dará lugar a recomendaciones diferentes para subgrupos distintos, entonces puede ser mejor presentar solamente los resultados generales del análisis.

- ¿Existen diferencias estadísticamente significativas entre los subgrupos?

Para establecer si hay un efecto diferente de una intervención en diferentes situaciones, las magnitudes de los efectos en diferentes subgrupos se deben comparar directamente entre sí. En particular, no se debe comparar la significación estadística de los resultados dentro de análisis de subgrupos separados. Ver sección 9.6.3.1.

- ¿Los análisis examinan las relaciones intra o entre estudios?

Las diferencias en los subgrupos según características del paciente y la intervención que se observan dentro de los estudios son más fiables que los análisis de subconjuntos de estudios. Si dichas relaciones entre estudios se replican entre los estudios entonces se reafirma la confianza en los hallazgos.

9.6.7 Investigación del efecto del riesgo basal

Una fuente potencialmente importante de heterogeneidad entre una serie de estudios es cuando el riesgo promedio subyacente del suceso de resultado varía entre los estudios. El riesgo basal de un suceso particular se puede considerar como una medida agregada de los factores de *case-mix* como la edad o la gravedad de la enfermedad. Generalmente se mide como el riesgo del suceso observado en el grupo control para cada estudio (el riesgo del grupo control [RGC]). La definición es controvertida en cuanto a su relevancia para la práctica clínica porque el riesgo basal representa un resumen de factores de riesgo conocidos y no conocidos. También aparecen problemas porque el riesgo basal dependerá de la duración del seguimiento, que a menudo varía entre los estudios. Sin embargo, el riesgo basal ha recibido atención particular en el metanálisis porque la información está fácilmente disponible una vez

que los datos dicotómicos se preparan para utilizarlos en los metanálisis. Sharp proporciona una discusión completa del tema (Sharp 2000).

La intuición sugeriría que es más o menos probable que los participantes se beneficien de una intervención efectiva según su estado de riesgo. Sin embargo, la relación entre el riesgo basal y el efecto de la intervención es un tema complicado. Por ejemplo, suponga que una intervención es igualmente beneficiosa porque para todos los pacientes reduce el riesgo de un accidente cerebrovascular al 80% del riesgo basal. Entonces la misma no es igualmente beneficiosa en cuanto a las diferencias absolutas de riesgos, ya que reduce una tasa de accidente cerebrovascular del 50% en 10 puntos porcentuales hasta alcanzer el 40% (número necesario a tratar = 10), pero reduce una tasa de accidente cerebrovascular del 20% en 4 puntos porcentuales hasta situarla en el 16% (número necesario a tratar = 25).

Diferentes estadísticos resumen (razón de riesgos, odds ratio y diferencia de riesgos) presentarán diferentes relaciones con el riesgo basal. Generalmente se prefiere utilizar en el metanálisis los estadísticos resumen que tengan la menor relación con el riesgo basal (ver Sección 9.4.4.4).

La investigación de cualquier relación entre las estimaciones del efecto y el riesgo del grupo control también se complica por un fenómeno técnico conocido como regresión a la media. El mismo surge porque el riesgo del grupo control forma parte integral de la estimación del efecto. Un alto riesgo en el grupo control, observado completamente por azar, dará lugar como promedio a una estimación del efecto más alta de la esperada, y *viceversa*. Este fenómeno provoca una falsa correlación entre las estimaciones del efecto y los riesgos del grupo control. Hay métodos disponibles, aunque requieren programas informáticos sofisticados, para corregir la regresión a la media (McIntosh 1996, Thompson 1997). Los mismos se deberían utilizar para dichos análisis y se recomienda experiencia estadística.

9.6.8 Análisis de dosis respuesta

Los principios de la metarregresión se pueden aplicar a las relaciones entre el efecto de la intervención y la dosis (llamada frecuentemente dosis respuesta), la intensidad del tratamiento o la duración del tratamiento (Greenland 1992, Berlin 1993). Las conclusiones acerca de diferencias en el efecto debido a diferencia en la dosis (o factores similares) tienen un fundamento sólido si los participantes se asignan al azar a una u otra dosis dentro de un estudio y se encuentra una relación consistente entre estudios similares. Aunque los revisores deben considerar estos efectos, particularmente como una posible explicación de la heterogeneidad, deben tener precaución con respecto a establecer conclusiones basadas en las diferencias entre estudios. Los revisores deben tener particular precaución cuando planteen que no existe una relación dosis respuesta, debido a la baja potencia estadística de muchos análisis de metarregresión para detectar verdaderas relaciones.

9.7 Análisis de sensibilidad

El proceso de realizar una revisión sistemática incluye una secuencia de decisiones. Aunque muchas de estas decisiones son claramente objetivas y no contenciosas, algunas serán de alguna manera arbitrarias o poco claras. Por ejemplo, si entre los criterios de inclusión hay un valor numérico, habitualmente la selección del valor es arbitraria: por ejemplo, la definición de grupos de personas con edad avanzada puede tener razonablemente límites inferiores como 60, 65, 70 ó 75 años, o cualquier valor entre ellos. Otras decisiones pueden ser poco claras porque el informe de un estudio no pudo incluir la información necesaria. Algunas decisiones son poco claras porque incluyen estudios que por sí mismos nunca obtuvieron la información requerida: por ejemplo, los resultados de los que desafortunadamente se perdieron durante el seguimiento. Otras decisiones son poco claras porque no hay consenso sobre el mejor método estadístico a utilizar en un problema particular.

Es aconsejable comprobar si los hallazgos de una revisión sistemática no dependen de decisiones arbitrarias o poco claras. Un análisis de sensibilidad es una repetición del análisis o el metanálisis primario, donde decisiones o rangos de valores alternativos sustituyen decisiones que fueron arbitrarias o poco claras. Por ejemplo, si la elegibilidad de algunos estudios en el metanálisis es dudosa debido a que no contienen todos los detalles, el análisis de sensibilidad puede conllevar realizar dos veces el metanálisis: primero, con la inclusión de todos los estudios y segundo, con la inclusión de los que son claramente elegibles. Un análisis de sensibilidad responde la pregunta "¿Los hallazgos son sólidos con respecto a las decisiones tomadas en el proceso de obtenerlos?".

Hay muchos puntos de decisión dentro del proceso de una revisión sistemática que pueden generar la necesidad de realizar un análisis de sensibilidad. Por ejemplo,

Búsqueda de los estudios:

- ¿Se deben incluir en la revisión los resúmenes cuyos resultados no se pueden confirmar en publicaciones posteriores?

Criterios de elegibilidad:

- Características de los participantes: cuando la mayoría pero no todas las personas en un estudio cumplieron un rango de edad, ¿se debe incluir el estudio?
- Características de la intervención: ¿qué rango de dosis se debe incluir en un metanálisis?
- Características del comparador: ¿qué criterios se requieren para definir la atención habitual que se utilizará como grupo comparador?
- Características del resultado: ¿qué punto temporal o rango de puntos temporales son elegibles para inclusión?
- Diseño del estudio: ¿se debe incluir la evaluación cegada y no cegada del resultado, o la inclusión del estudio debe estar limitada por otros aspectos metodológicos?

¿Qué datos se deben analizar?

- Datos de tiempo hasta el suceso: ¿qué suposiciones se deben hacer con respecto a la distribución de los datos censurados?
- Datos continuos: cuando faltan las desviaciones estándar, ¿cuándo y cómo se deben imputar? ¿Los análisis se deben basar en las puntuaciones de cambio o en los valores finales?
- Escalas ordinales: ¿qué puntos de corte se deben utilizar para dicotomizar las escalas ordinales cortas en dos grupos?
- Ensayos aleatorizados por conglomerados: ¿qué valores del coeficiente de correlación intraclase se deben utilizar cuando los análisis del ensayo no se han ajustado para el agrupamiento?
- Ensayos cruzados: ¿qué valores del coeficiente de correlación dentro de los sujetos se deben utilizar cuando el mismo no está disponible en los informes primarios?
- Todos los análisis: ¿qué suposiciones se deben hacer acerca de los resultados faltantes para facilitar los análisis por intención de tratar? ¿Se deben utilizar las estimaciones ajustadas o las no ajustadas?

Métodos de análisis:

- Para el análisis, ¿se deben utilizar los métodos de efectos fijos o los de efectos aleatorios?
- Para los resultados dicotómicos, ¿se deben utilizar los odds ratios, las razones de riesgos o las diferencias de riesgos?
- Y para los resultados continuos, cuando varias escalas hayan evaluado la misma dimensión, ¿se deben analizar los resultados como diferencias de medias estandarizada entre todas las escalas, o como diferencias de medias individualmente para cada escala?

Algunos análisis de sensibilidad se pueden preespecificar en el protocolo del estudio, pero muchos aspectos adecuados para el análisis de sensibilidad sólo se identifican durante el proceso de revisión, cuando se identifican las particularidades individuales de los estudios bajo investigación. Cuando los análisis de sensibilidad muestran que el resultado y las

conclusiones generales no se ven afectados por las diferentes decisiones que se pudieron tomar durante el proceso de revisión, los resultados de una revisión se pueden considerar como de alto grado de certidumbre. Cuando los análisis de sensibilidad identifican decisiones particulares o información faltante que influye marcadamente en los hallazgos de la revisión, se pueden desplegar mayores cantidades de recursos para tratar de resolver las incertidumbres y obtener información adicional, posiblemente mediante el contacto con los autores del ensayo, y obtener datos de pacientes individuales. Si no es posible lograrlo, los resultados se deben interpretar con un grado apropiado de precaución. Estos hallazgos pueden generar propuestas para investigaciones adicionales futuras.

El informe de los análisis de sensibilidad en una revisión sistemática se puede realizar mejor mediante la creación de una tabla resumen. En pocas ocasiones ésta es suficientemente informativa para producir diagramas de árbol individuales para cada análisis de sensibilidad realizado.

En ocasiones los análisis de sensibilidad se confunden con los análisis de subgrupos. Aunque algunos análisis de sensibilidad limitan el análisis a un subconjunto del total de estudios, los dos métodos difieren en dos aspectos. Primero, los análisis de sensibilidad no intentan calcular el efecto de la intervención en el grupo de estudios eliminados del análisis, mientras que en los análisis de subgrupos se producen estimaciones para cada subgrupo. Segundo, en los análisis de sensibilidad se hacen comparaciones informales entre diferentes formas de calcular lo mismo, mientras que en los análisis de subgrupos se hacen comparaciones estadísticas formales entre los subgrupos.

9.8 Información del capítulo

Editores: Jonathan J Deeks, Julian PT Higgins y Douglas G Altman en nombre del Grupo Cochrane de Métodos Estadísticos.

La versión en inglés de este capítulo se debe citar como: Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Autores que contribuyeron: Doug Altman, Deborah Ashby, Jacqueline Birks, Michael Borenstein, Marion Campbell, Jon Deeks, Matthias Egger, Julian Higgins, Joseph Lau, Keith O'Rourke, Rob Scholten, Jonathan Sterne, Simon Thompson y Anne Whitehead.

Agradecimientos: Se agradece a las siguientes personas por sus útiles comentarios sobre los primeros borradores: Bodil Als-Nielsen, Doug Altman, Deborah Ashby, Jesse Berlin, Joseph Beyene, Jacqueline Birks, Michael Bracken, Marion Campbell, Chris Cates, Wendong Chen, Mike Clarke, Albert Cobos, Esther Coren, Francois Curtin, Roberto D'Amico, Keith Dear, Jon Deeks, Heather Dickinson, Diana Elbourne, Simon Gates, Paul Glasziou, Christian Gluud, Peter Herbison, Julian Higgins, Sally Hollis, David Jones, Steff Lewis, Philippa Middleton, Nathan Pace, Craig Ramsey, Keith O'Rourke, Rob Scholten, Guido Schwarzer, Jack Sinclair, Jonathan Sterne, Simon Thompson, Andy Vail, Clarine van Oel, Paula Williamson y Fred Wolf.

Recuadro 9.8.a: El Grupo Cochrane de Métodos Estadísticos

Los temas estadísticos son un aspecto fundamental de una gran parte del trabajo de la Colaboración Cochrane. El Grupo de Métodos Estadísticos (GME) es un forum donde se discuten todos los aspectos estadísticos relacionados con el trabajo de la Colaboración Cochrane. Tiene un alcance amplio y cubre los aspectos relacionados con los métodos estadísticos, el entrenamiento, el programa informático y la investigación. También intenta asegurar que haya un apoyo estadístico y técnico adecuado para los grupos de revisión.
El GME data de 1993. Actualmente la membresía del GME parte de la membresía de la lista de discusión por correo electrónico del grupo. Esta lista se utiliza para discutir todos los aspectos de importancia para el grupo como la investigación, el entrenamiento, el programa informático o la administración. El grupo tiene cerca de 130 miembros de alrededor de 20 países. Se alienta a todos los estadísticos que trabajan con los Grupos Cochrane de Revisión (GCR) a que se unan al GME.

Específicamente, los objetivos del grupo son:
Desarrollar una política de asesoramiento general para la Colaboración en todos los aspectos estadísticos relevantes para las revisiones sistemáticas de intervenciones sanitarias.
Responsabilizarse por los capítulos de este *Manual* relacionados con la estadística.
Coordinar el apoyo estadístico práctico a los GCR.
Realizar talleres de entrenamiento y talleres sobre temas emergentes, de ser necesario.
Contribuir y revisar el contenido estadístico de los materiales de entrenamiento proporcionados dentro de la Colaboración.
Desarrollar y validar el programa estadístico utilizado dentro de la Colaboración.
Generar y mantener actualizado el Grupo de Métodos Estadísticos, detallar sus áreas de interés y experiencia y mantener una lista de discusión por correo electrónico como un forum para discutir aspectos metodológicos importantes.
Mantener una agenda de investigación regida por aspectos importantes para el funcionamiento actual y futuro de la colaboración, y estimular la investigación que se ajuste a la agenda.

Web site: www.cochrane-smg.org

9.9 Referencias

Adams 2005
Adams NP, Bestall JB, Malouf R, Lasserson TJ, Jones PW. Beclomethasone versus placebo for chronic asthma. *Cochrane Database of Systematic Reviews* 2005, Issue 1. Art No: CD002738.

Agresti 1996
Agresti A. *An introduction to categorical data analysis*. New York (NY): John Wiley & Sons, 1996.

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: 240-248.

Altman 1996
Altman DG, Bland JM. Detecting skewness from summary information. *BMJ* 1996; 313: 1200-1200.

Berlin 1993
Berlin JA, Longnecker MP, Greenland S. Meta-analysis of epidemiologic dose-response data. *Epidemiology* 1993; 4: 218-228.

Berlin 1994
Berlin JA, Antman EM. Advantages and limitations of metaanalytic regressions of clinical trials data. *Online Journal of Current Clinical Trials* 1994; Doc No 134.

Berlin 2002
Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman KA, Anti-Lymphocyte Antibody Induction Therapy Study Group. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Statistics in Medicine* 2002; 21: 371-387.

Bradburn 2007
Bradburn MJ, Deeks JJ, Berlin JA, Russell LA. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine* 2007; 26: 53-77.

Chinn 2000
Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine* 2000; 19: 3127-3131.

Cooper 1980
Cooper HM, Rosenthal R. Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin* 1980; 87: 442-449.

Crawford 2007
Crawford F, Hollis S. Topical treatments for fungal infections of the skin and nails of the feet. *Cochrane Database of Systematic Reviews* 2007, Issue 3. Art No: CD001434.

Deeks 1998
Deeks JJ. Systematic reviews of published evidence: Miracles or minefields? *Annals of Oncology* 1998; 9: 703-709.

Deeks 2001
Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG (editors). *Systematic Reviews in Health Care: Meta-analysis in Context* (2nd edition). London (UK): BMJ Publication Group, 2001.

Deeks 2002
Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* 2002; 21: 1575-1600.

DerSimonian 1986
DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; 7: 177-188.

Egger 1997
Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315: 629-634.

Engels 2000
Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Statistics in Medicine* 2000; 19: 1707-1728.

Greenland 1985
Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. *Biometrics* 1985; 41: 55-68.

Greenland 1987
Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiologic Reviews* 1987; 9: 1-30.

Greenland 1992
Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *American Journal of Epidemiology* 1992; 135: 1301-1309.

Hasselblad 1995
Hasselblad VIC, McCrory DC. Meta-analytic tools for medical decision making: A practical guide. *Medical Decision Making* 1995; 15: 81-96.

Higgins 2002
Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002; 21: 1539-1558.

Higgins 2003
Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557-560.

Higgins 2004
Higgins JPT, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine* 2004; 23: 1663-1682.

Higgins 2008a
Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society Series A* (in press, 2008).

Higgins 2008b
Higgins JPT, White IR, Anzures-Cabrera J. Meta-analysis of skewed data: combining results reported on log-transformed or raw scales. *Statistics in Medicine* (in press, 2008).

Kjaergard 2001
Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Annals of Internal Medicine* 2001; 135: 982-989.

Laupacis 1988
Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *New England Journal of Medicine* 1988; 318: 1728-1733.

Mantel 1959
Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 1959; 22: 719-748.

McIntosh 1996
McIntosh MW. The population risk as an explanatory variable in research synthesis of clinical trials. *Statistics in Medicine* 1996; 15: 1713-1728.

Moher 2005
Moher M, Hey K, Lancaster T. Workplace interventions for smoking cessation. *Cochrane Database of Systematic Reviews* 2005, Issue 2. Art No: CD003440.

Morgenstern 1982
Morgenstern H. Uses of ecologic analysis in epidemiologic research. *American Journal of Public Health* 1982; 72: 1336-1344.

O'Rourke 1989
O'Rourke K, Detsky AS. Meta-analysis in medical research: strong encouragement for higher quality in individual research efforts. *Journal of Clinical Epidemiology* 1989; 42: 1021-1026.

Oxman 1992
Oxman AD, Guyatt GH. A consumers guide to subgroup analyses. *Annals of Internal Medicine* 1992; 116: 78-84.

Pittler 2003
Pittler MH, Ernst E. Kava extract versus placebo for treating anxiety. *Cochrane Database of Systematic Reviews* 2003, Issue 1. Art No: CD003383.

Poole 1999
Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *American Journal of Epidemiology* 1999; 150: 469-475.

Sackett 1996
Sackett DL, Deeks JJ, Altman DG. Down with odds ratios! *Evidence Based Medicine* 1996; 1: 164-166.

Sackett 1997
Sackett DL, Richardson WS, Rosenberg W, Haynes BR. *Evidence-Based Medicine: How to Practice and Teach EBM*. Edinburgh (UK): Churchill Livingstone, 1997.

Sharp 2000
Sharp SJ. Analysing the relationship between treatment benefit and underlying risk: precautions and practical recommendations. In: Egger M, Davey Smith G, Altman DG (editors). *Systematic Reviews in Health Care: Meta-analysis in Context* (2nd edition). London (UK): BMJ Publication Group, 2000.

Sinclair 1994
Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. *Journal of Clinical Epidemiology* 1994; 47: 881-889.

Thompson 1997
Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. *Statistics in Medicine* 1997; 16: 2741-2758.

Thompson 1999
Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; 18: 2693-2708.

Thompson 2002
Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* 2002; 21: 1559-1574.

Whitehead 1991
Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomised clinical trials. *Statistics in Medicine* 1991; 10: 1665-1677.

Whitehead 1994
Whitehead A, Jones NMB. A meta-analysis of clinical trials involving different classifications of response into ordered categories. *Statistics in Medicine* 1994; 13: 2503-2515.

Yusuf 1985
Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomised trials. *Progress in Cardiovascular Diseases* 1985; 27: 335-371.

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: 93-98.

ÍNDICE
CAPÍTULO 10: ANÁLISIS DEL SESGO DE INFORME

PUNTOS CLAVE

10.1 INTRODUCCIÓN

TABLA 10.1.A: Definiciones de algunos tipos de sesgos de informe

10.2 TIPOS DE SESGOS DE INFORME Y LAS PRUEBAS QUE LOS APOYAN

10.2.1 Sesgo de publicación
Figura 10.2.a: Publicación completa acumulativa de los resultados presentados inicialmente como resúmenes de 45 estudios que reportaron el tiempo hasta la publicación que siguió a la investigación presentada en reuniones y congresos
Taula 10.2.a: Estado de publicación de cinco cohortes de proyectos de investigación aprobados por los comités de ética o las comisiones de investigación institucionales que habían finalizado y se habían analizado en el momento del seguimiento (Adaptado de Hopewell y cols. ([Hopewell 2008])
Figura 10.2.b: Sesgo de publicación en ensayos clínicos debido a la significación estadística o la dirección de los resultados del ensayo (Adaptado de Hopewell y cols. [Hopewell 2008]).
10.2.2.1 Sesgo de lapso de tiempo
10.2.2.2 ¿Quién es responsable del sesgo de publicación?
10.2.2.3 La influencia de la financiación externa y los intereses comerciales
10.2.2 Otros sesgos de informe
10.2.2.1 Sesgo de publicación duplicada (múltiple)
10.2.2.2 Sesgo de ubicación
10.2.2.3 Sesgo de citación
10.2.2.4 Sesgo de idioma
10.2.2.5 Sesgo de informe de resultado

10.3 CÓMO EVITAR LOS SESGOS DE INFORME

10.3.1 Implicaciones de las pruebas relacionadas con los sesgos de informe
10.3.2 Inclusión de estudios no publicados en las revisiones sistemáticas
10.3.3 Registros de ensayos y sesgo de publicación

10.4 DETECCIÓN DE LOS SESGOS DE INFORME

10.4.1 Gráficos de embudo
Figura 10.4.a: Gráficos de embudo hipotéticos
10.4.2 Diferentes motivos para la asimetría del gráfico de embudo
Tabla 10.4.a: Posibles fuentes de asimetría en los gráficos de embudo
Figura 10.4.b: Gráficos de embudo de contorno mejorado
10.4.3 Pruebas para la asimetría del gráfico de embudo
Tabla 10.4.b: Pruebas propuestas para analizar la asimetría del gráfico de embudo

10.4.3.1 Recomendaciones para las pruebas de asimetría del gráfico de embudo para todo tipo de resultado
10.4.4 Análisis de sensibilidad
10.4.4.1 Comparación de las estimaciones de efectos fijos y aleatorios
Figura 10.4.c: Comparación del efecto del magnesio intravenoso sobre la mortalidad tras un infarto de miocardio mediante estimaciones metanalíticas del modelo de efectos fijos de efectos aleatorios
10.4.4.2 Recortar y rellenar
10.4.4.3 N a prueba de fallos
10.4.4.4 Otros modelos de selección
10.4.4.5 Análisis de sensibilidad según los modelos de selección
10.4.4.6 Pruebas para el exceso de estudios con resultados significativos
10.4.5 Resumen

10.5 INFORMACIÓN DEL CAPÍTULO

Recuadro 10.5.a: El Grupo Cochrane de Métodos de Sesgo

10.6 REFERENCIAS

CAPÍTULO 10: ANÁLISIS DEL SESGO DE INFORME

Editores: Jonathan AC Sterne, Matthias Egger y David Moher del Grupo Cochrane de Métodos de Sesgo.

Puntos clave

- Solamente una parte de los proyectos de investigación se publicarán en fuentes fácilmente identificables por los autores de las revisiones sistemáticas. Los sesgos de informe surgen cuando la diseminación de los hallazgos de la investigación están influenciados por la naturaleza y la dirección de los resultados.
- La contribución hecha al total de las pruebas en las revisiones sistemáticas por los estudios con resultados estadísticamente no significativos es tan importante como la de los estudios con resultados estadísticamente significativos.
- Las pruebas convincentes de la presencia de varios tipos de sesgos de resultado (esbozados en este capítulo) demuestran la necesidad de buscar de forma exhaustiva los estudios que cumplan los criterios de elegibilidad para una revisión Cochrane.
- El registro prospectivo de los ensayos, que es ahora un requerimiento para la publicación en muchas revistas, tiene el potencial de reducir significativamente los efectos del sesgo de publicación.
- Los gráficos de embudo se pueden utilizar para las revisiones con un número suficiente de estudios incluidos, pero no se debe considerar que un gráfico de embudo asimétrico equivale a la presencia de sesgo de publicación.
- Existen varios métodos disponibles para probar la asimetría de un grafico de embudo y en el capítulo se incluyen recomendaciones para seleccionar una prueba apropiada.

10.1 Introducción

La diseminación de los hallazgos de las investigaciones no se divide en publicadas y no publicadas, sino que es un proceso continuo que varía desde compartir el borrador de los artículos entre colegas, pasa por presentaciones en reuniones y resúmenes publicados, hasta artículos en revistas que están indexadas en las principales bases de datos bibliográficas (Smith 1999). Se ha reconocido ampliamente que solamente una parte de los proyectos de investigación logran finalmente ser publicados en una revista indexada, por lo que son fácilmente identificables para las revisiones sistemáticas.

Los sesgos de informe surgen cuando la diseminación de los hallazgos de las investigaciones está influenciada por la naturaleza y la dirección de los resultados. Es más probable que se publiquen los resultados "positivos" estadísticamente significativos que indican que una intervención funciona, más probable que se publiquen más rápido, más probable que se publiquen en inglés, más probable que se publiquen más de una vez, más probable que se publiquen en revistas de gran impacto y, relacionado con este último punto, y más probable que se citen por otros. La contribución hecha al total de las pruebas en las revisiones sistemáticas por los estudios con resultados estadísticamente no significativos es tan importante como la de los estudios con resultados estadísticamente significativos

La Tabla 10.1.a resume algunos tipos diferentes de sesgo de informe. Los mismos se consideran en más detalle en la Sección 10.2, y se destacan en particular las pruebas que apoyan la presencia de cada sesgo. En la Sección 10.3 se discuten los enfoques para evitar los sesgos de informe en las revisiones Cochrane, y en la Sección 10.4 se abordan los gráficos de embudo y los métodos estadísticos para detectar los sesgos potenciales. Aunque para los fines de la discusión de estos sesgos en ocasiones los resultados estadísticamente significativos ($p < 0,05$) se denotarán como resultados "positivos" y los resultados estadísticamente no significativos o nulos como resultados "negativos", los revisores Cochrane no deben utilizar tales denominaciones.

Tabla 10.1.a: Definiciones de algunos tipos de sesgos de informe

Tipo de sesgo de informe	Definición
Sesgo de publicación	La *publicación* o *no publicación* de los hallazgos de la investigación, según la naturaleza y la dirección de los resultados
Sesgo de lapso de tiempo	La publicación *rápida* o *demorada* de los hallazgos de la investigación, según la naturaleza y la dirección de los resultados
Sesgo de publicación múltiple (duplicada)	La publicación *múltiple* o *única* de los hallazgos de la investigación, según la naturaleza y la dirección de los resultados
Sesgo de ubicación	La publicación de los hallazgos de la investigación en revistas con *facilidad de acceso* o *niveles de indexación* en bases de datos estándar, según la naturaleza y la dirección de los resultados
Sesgo de citación	La *citación* o *no citación* de los hallazgos de la investigación, según la naturaleza y la dirección de los resultados
Sesgo de idioma	La publicación de los hallazgos de la investigación *en un idioma particular*, según la naturaleza y la dirección de los resultados
Sesgo de informe de resultado	El *informe selectivo* de algunos resultados pero no de otros, según la naturaleza y la dirección de los resultados

10.2 Tipos de sesgos de informe y las pruebas que los apoyan

10.2.1 Sesgo de publicación

En un artículo de 1979, "El 'problema del cajón de archivos y la tolerancia a los resultados nulos" (The 'file drawer problem' and tolerance for null results), Rosenthal describió un escenario lóbrego donde "las revistas se llenan con el 5% de los estudios que muestran errores Tipo I, mientras que los cajones de archivos de los laboratorios están llenos con el 95% de los estudios que muestran resultados no significativos (p.ej. p > 0,05)" (Rosenthal 1979). El problema de los cajones de archivos se ha sospechado durante mucho tiempo en las ciencias sociales: una revisión de las revistas de psicología encontró que de 294 estudios publicados en los años 50, el 97,3% rechazó la hipótesis nula al nivel del 5% (p < 0,05) (Sterling 1959). El estudio se actualizó y se complementó con otras tres revistas (*New England Journal of Medicine*, *American Journal of Epidemiology, American Journal of Public Health*) (Sterling 1995). Poco ha cambiado en las revistas de psicología (el 95,6% informó resultados significativos) y en las revistas médicas generales y de salud pública también se encontró una alta proporción de resultados estadísticamente significativos (85,4%). Se han informado resultados similares en muchas áreas diferentes como la medicina de urgencia (Moscati 1994), la medicina alternativa y complementaria (Vickers 1998, Pitter 2000) y en los ensayos de accidente cerebrovascular agudo (Liebeskind 2006).

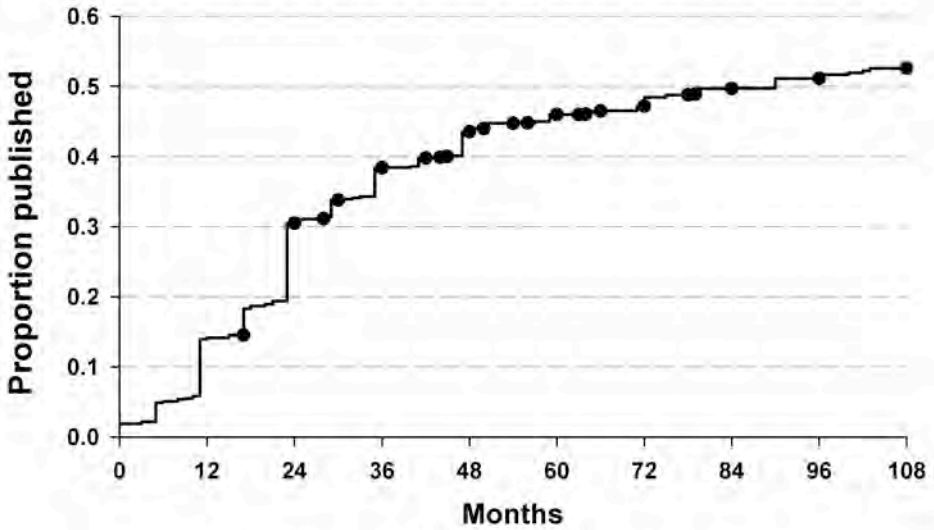
Es posible que los estudios que indican un efecto beneficioso de la intervención o un tamaño mayor del efecto se publiquen, mientras que una cantidad similar de datos que apuntan en la otra dirección permanezcan sin publicarse. En esta situación, una revisión sistemática de los estudios publicados podría identificar un efecto beneficioso espurio de la intervención, o perderse un efecto adverso importante de una intervención. En la medicina cardiovascular, los investigadores que en 1980 encontraron un aumento en la tasa de muerte entre los pacientes con infarto agudo de miocardio tratados con un antiarrítmico clase 1 lo descartaron como debido al azar y no publicaron su ensayo en ese momento (Cowley 1993). Sus hallazgos podrían haber contribuido a detectar a tiempo el aumento de la mortalidad que luego se conoció que se asocia con el uso de los agentes antiarrítmicos clase 1 (Teo 1993, CLASP Collaborative Group 1994).

Los estudios que examinan empíricamente la presencia de sesgo de publicación se pueden ubicar en dos categorías: pruebas indirectas y directas. Los análisis de resultados publicados, como los que se describen anteriormente, pueden proporcionar solamente pruebas indirectas del sesgo de publicación, ya que se desconoce la proporción de todas las hipótesis probadas para las cuales la hipótesis nula es verdaderamente falsa. También existen pruebas directas significativas del sesgo de publicación. Roberta Scherer y cols. actualizaron recientemente una revisión sistemática que resume 79 estudios que describen la publicación completa posterior de investigaciones que inicialmente se presentaron como resumen o como un informe corto (Scherer 2007). Los datos de 45 estudios que incluyeron datos sobre el tiempo hasta la publicación se resumen en la Figura 10.2.a. Solamente cerca de la mitad de los resúmenes presentados en congresos se publicaron completos posteriormente (63% para los ensayos aleatorios) y la publicación posterior se asoció con resultados positivos (Scherer 2007).

Existen pruebas directas adicionales de varios estudios de cohortes de proyectos enviados a los comités de ética y a las comisiones de investigación institucionales (Easterbrook 1991, Dickersin 1992, Stern 1997, Deculier 2005, Decullier 2007), ensayos enviados a las autoridades que expiden licencias (Bardy 1998), análisis de registros de ensayos (Simes 1987) o de cohortes de ensayos financiados por agencias de financiación específicas (Dickersin 1993). Para cada cohorte de proyectos de investigación se estableció contacto con los investigadores principales varios años después para determinar el estado de la publicación de cada estudio finalizado. Fue más probable que todos estos estudios se hubieran publicado si los efectos de la intervención eran grandes y estadísticamente significativos.

Hopewell y cols. finalizaron recientemente una revisión metodológica de dichos estudios, limitada a los que analizaron los ensayos clínicos por separado (Hopewell 2008). El porcentaje de publicación completa como artículo de revista en los cinco estudios incluidos en la revisión varió del 36% al 94% (Tabla 10.2.a). Fue consistentemente más probable que los resultados positivos se publicaran, comparados con los resultados negativos; las probabilidades de publicación fueron aproximadamente cuatro veces mayores si los resultados eran estadísticamente significativos (OR 3,90; IC del 95%: 2,68 a 5,68) como se muestra en la Figura 10.2.b. Otros factores como el tamaño del estudio, la fuente de financiación y el rango académico y el sexo del investigador principal no se asociaron de manera consistente con la probabilidad de publicación o no fue posible evaluarlos por separado en los ensayos clínicos (Hopewell 2008).

Figura 10.2.a: Publicación completa acumulativa de los resultados presentados inicialmente como resúmenes de 45 estudios que reportaron el tiempo hasta la publicación que siguió a la investigación presentada en reuniones y congresos



month	0	12	24	36	48	60	72	84	96	108
# published	362	2,460	3,348	1,519	800	280	282	84	27	10
# remaining	20,227	19,091	16,313	10,758	9,032	6,518	4,030	1,803	1,352	246

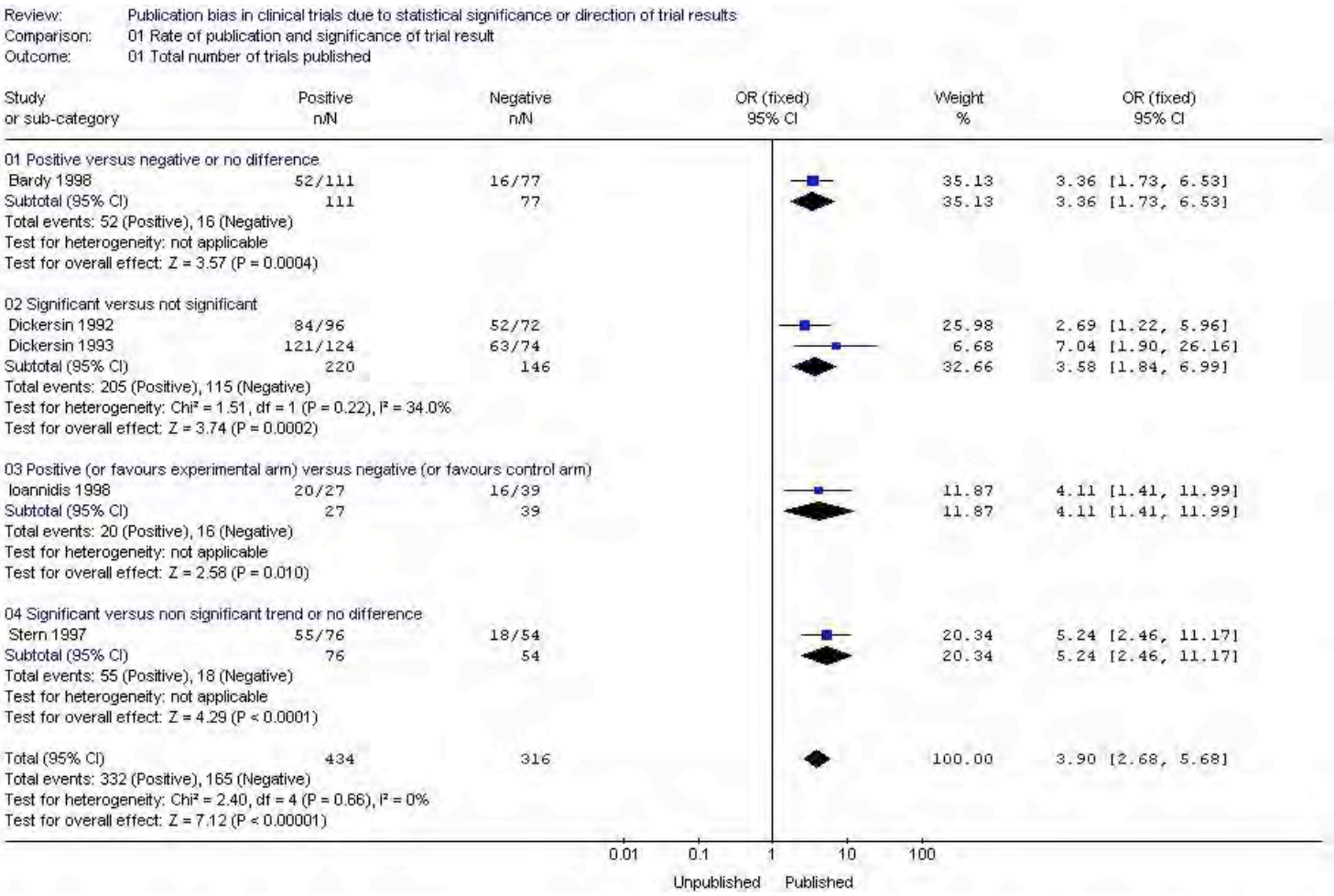
N = 20,227 abstracts
Circles show points where data censored because reports stopped follow-up.

Tabla 10.2.a: Estado de publicación de cinco cohortes de proyectos de investigación aprobados por los comités de ética o las comisiones de investigación institucionales que habían finalizado y se habían analizado en el momento del seguimiento (Adaptado de Hopewell y cols. ([Hopewell 2008]).

	Universidad Johns Hopkins, Baltimore	Institutos Nacionales de Salud, EE.UU.	Hospital Royal Prince Alfred, Sydney	Agencia Nacional para la Medicina, Finlandia	Institutos Nacionales de Salud, EE.UU., Ensayos multicéntricos en VIH/SIDA
Referencia	Dickersin 1992	Dickersin 1993	Stern 1997	Bardy 1998	Ioannidis 1998
Período de aprobación	1980	1979	1979-88	1987	1986-1996
Año de seguimiento	1988	1988	1992	1995	1996
Número de aprobados	168	198	130	188	66
Publicados	136 (81%)	184 (93%)	73 (56%)	68 (36%)	36 (54%)
Positivos*	84/96 (87%)	121/124 (98%)	55/76 (72%)	52/111 (47%)	20/27 (75%)
Negativos*	52/72 (72%)	63/74 (85%)	3/15 (20%)	5/44 (11%)	16/39 (41%)
No concluyentes/ nulos (si se evaluaron por separado)	No evaluados	No evaluados	15/39 (38%)	11/33 (33%)	No evaluados

* Las definiciones variaron en los estudios.

Figura 10.2.b: Sesgo de publicación en ensayos clínicos debido a la significación estadística o la dirección de los resultados del ensayo (Adaptado de Hopewell y cols. [Hopewell 2008]).



10.2.1.1 Sesgo de lapso de tiempo

Aún en la actualidad los estudios aparecen impresos muchos años después de su aprobación por los comités de ética. Hopewell y cols. revisaron los estudios que examinaron el tiempo hasta las publicaciones de los resultados de los ensayos clínicos (Hopewell 2007a). Los dos estudios incluidos en esta revisión (Stern 1997, Ioannidis 1998) encontraron que cerca de la mitad de todos los ensayos se publicaron, y que los que tenían resultados positivos se publicaron como promedio de dos a tres años antes que los que tenían resultados negativos o nulos.

Entre los proyectos enviados al Comité de Ética del Royal Prince Alfred Hospital en Sydney, Australia, aproximadamente el 85% de los estudios con resultados significativos, comparados con el 65% de los estudios con resultados nulos, se habían publicado después de 10 años (Stern 1997). La mediana del tiempo hasta la publicación fue 4,7 años para los estudios con resultados significativos y 8,0 años para los estudios con resultados negativos/nulos. De manera similar, los ensayos realizados por grupos de ensayos muticéntricos en el campo de la infección por VIH/SIDA en los Estados Unidos aparecieron como promedio 4,3 años después del inicio del reclutamiento de los pacientes si los resultados eran estadísticamente significativos, pero se demoraron 6,5 años en publicarse si los resultados eran negativos (Ioannidis 1998). Un estudio reciente ha encontrado resultados similares (Decullier 2005). El hecho de que una proporción importante de estudios permanezca sin publicar incluso una década después de que el estudio se completó y analizó es problemático debido a que información potencialmente importante permanece fuera del alcance de los autores de revisiones sistemáticas y los consumidores.

Ioannidis y cols. también encontraron que los ensayos con resultados positivos y negativos difirieron poco en el tiempo que demoraron hasta el seguimiento completo (Ioannidis 1998). En su lugar, el lapso de tiempo se atribuyó a las diferencias en el tiempo desde la finalización hasta la publicación (Ioannidis 1998). Estos hallazgos indican que es posible introducir el sesgo de lapso de tiempo en las revisiones sistemáticas incluso en situaciones donde la mayoría o todos los estudios finalmente se publicarán. Los estudios con resultados positivos dominarán la bibliografía e introducirán sesgo durante muchos años hasta que los resultados negativos, pero igualmente importantes, finalmente aparezcan. Además, es probable que se encuentren eventos adversos poco frecuentes más tarde en el proceso de investigación, comparados con los efectos beneficiosos a corto plazo.

10.2.1.2 ¿Quién es responsable del sesgo de publicación?

Los estudios con resultados negativos pueden permanecer sin publicar porque los autores no redacten el texto ni lo envíen a las revistas, porque la revisión por pares de dichos estudios sea menos favorable o porque los editores sencillamente no deseen publicar resultados negativos. En ocasiones el proceso de revisión por pares es poco fiable y susceptible de subjetividad, sesgo y conflicto de intereses (Peters 1982, Godlee 1999). Estudios experimentales en los cuales se enviaron manuscritos de prueba a revisores pares o revistas mostraron que era más probable que los revisores pares hicieran una evaluación favorable si los resultados estaban de acuerdo con sus propios puntos de vista (Mahoney 1977, Epstein 1990, Ernst 1994). Por ejemplo, cuando a un grupo seleccionado de autores se le solicitó que fueran revisores pares de un artículo ficticio sobre estimulación nerviosa eléctrica transcutánea (ENET), estuvieron influenciados por sus propios hallazgos y presuposiciones. Otros estudios no han mostrado asociación entre la publicación de textos enviados y los resultados del estudio (Abbot 1998, Olson 2002), lo que indica que aunque los revisores pares pueden tener creencias firmes que influirán en sus evaluaciones, no hay un sesgo general a favor o en contra de los hallazgos positivos.

Varios estudios han preguntado directamente a los autores por qué no han publicado sus hallazgos. La respuesta más frecuente fue que los mismos no eran suficientemente interesantes como para merecer su publicación (p.ej. era poco probable que las revistas

aceptaran los textos) (Easterbrook 1991, Dickersin 1992, Stern 1997, Weber 1998, Decullier 2005) o los investigadores no tenían tiempo suficiente para preparar el texto (Weber 1998, hartling 2004). Pocas veces se mencionó el rechazo del texto por parte de una revista como un motivo para no publicarlo. Por lo tanto, el envío selectivo de los artículos por los autores, en lugar de la recomendación selectiva de revisores pares y la aceptación selectiva de los editores, parece ser el contribuyente fundamental al sesgo de publicación. Además, Dickersin y cols. examinaron el tiempo desde el envío del texto (a la revista *JAMA*) hasta la publicación completa y no encontraron asociación entre este tiempo y cualquier característica del estudio examinada, incluida la significación estadística de los resultados del estudio (Dickersin 2002). Por lo tanto, el sesgo de lapso de tiempo también puede ser resultado del envío retrasado de los textos para su publicación por parte de los autores, en lugar de la publicación demorada en las revistas.

10.2.1.3 La influencia de la financiación externa y los intereses comerciales

Se ha encontrado que la financiación externa se asocia con la publicación, independientemente de la significación estadística de los resultados (Dickersin 1997). La financiación por agencias gubernamentales se asoció significativamente con la publicación en tres cohortes de proyectos enviados a comités de ética (Easterbrook 1991, Dickersin 1992, Stern 1997), mientras que en dos estudios (Easterbrook 1991, Dickersin 1992) fue menos probable que los estudios patrocinados por la industria farmacéutica se publicaran. De hecho, una gran proporción de ensayos clínicos enviados por compañías farmacéuticas a las autoridades que otorgan licencias permanecen sin publicar (Hemminki 1980, Baedy 1998).

En una revisión sistemática Lexchin y cols. identificaron 30 estudios publicados entre 1966 y 2002 que examinaron si la financiación de los estudios de fármacos por la industria farmacéutica se asociaba con resultados favorables para los patrocinadores. Encontraron que era menos probable que se publicaran las investigaciones financiadas por las compañías farmacéuticas, comparadas con las investigaciones financiadas por otras fuentes, y que era más probable que los estudios patrocinados por las compañías farmacéuticas tuvieran resultados favorables a los patrocinadores comparados con los estudios financiados por otros (Lexchin 2003). Otros estudios posteriores han examinado estas asociaciones y encontraron resultados similares (Bhandari 2004, Heres 2006). Heres y cols., en un estudio de comparaciones directas de antipsicóticos, encontraron que el resultado general de los ensayos favoreció al fármaco fabricado por el patrocinador de la industria en el 90% de los estudios analizados, y que algunos estudios similares informaron conclusiones opuestas, cada una de ellas en apoyo del producto del patrocinador del estudio (Heres 2006).

La implicación resultante es que la industria farmacéutica tiende a rechazar la publicación de los estudios negativos que ha financiado. Por ejemplo, se retiró un informe de un ensayo que compara la bioequivalencia de productos genéricos y de marca de levotiroxina que no produjo los resultados deseados por el patrocinador del estudio, Boots Pharmaceuticals, porque esta empresa tomó acciones legales contra la universidad y los investigadores. Las acciones de Boots, descritas en detalle por uno de los editores de *JAMA*, Drummond Rennie (Rennie 1997), dieron lugar a que la publicación del artículo (Dong 1997) se demorara cerca de siete años. En una encuesta nacional de miembros de la facultad de ciencias de la vida en los Estados Unidos, el 20% informó que habían tenido demoras de más de seis meses en la publicación de su trabajo y los motivos para la no publicación incluyeron “demorar la diseminación de resultados no deseados” (Blumenthal 1997). Las demoras en la publicación se asociaron con la participación en la comercialización y las relaciones científicas universidad-industria, así como el sexo masculino y el mayor rango académico del investigador (Blumenthal 1997).

10.2.2 Otros sesgos de informe

Introducción

Mientras que el sesgo de publicación se ha reconocido y discutido ampliamente, otros factores pueden contribuir a la inclusión sesgada de estudios en los metanálisis. De hecho, entre los estudios publicados la probabilidad de identificar estudios relevantes para el metanálisis también está influenciada por sus resultados. Estos sesgos han recibido mucha menos atención que el sesgo de publicación, pero sus consecuencias pudieran tener una importancia similar.

10.2.2.1 Sesgo de publicación duplicada (múltiple)

En 1989, Gøtzsche encontró que entre 244 informes de ensayos que comparaban fármacos anti-inflamatorios no esteroideos en la artritis reumatoide, 44 (18%) fueron publicaciones múltiples redundantes que se superpusieron significativamente con un artículo publicado previamente. Veinte artículos se publicaron dos veces, diez se publicaron tres veces y uno se publicó cuatro veces (Gøtzsche 1989). La producción de publicaciones múltiples de estudios únicos puede dar lugar a sesgo de varias maneras (Huston 1996). Más importante aún, es más probable que los estudios con resultados significativos den lugar a publicaciones y presentaciones múltiples (Easterbrook 1991), lo cual hace que sea más probable que se encuentren y se incluyan en un metanálisis. No siempre es evidente que múltiples publicaciones provienen del mismo estudio, y por ello es posible que un conjunto de participantes de un estudio se incluya dos veces en un análisis. Por lo tanto, la inclusión de datos duplicados puede dar lugar a la sobrestimación de los efectos de la intervención, como se demostró en los ensayos de la eficacia del ondansetrón para prevenir las nauseas y vómitos post-operatorios (Tramèr 1997).

Otros autores han descrito las dificultades y la frustración provocadas por la redundancia y la “desagregación” de las investigaciones médicas cuando los resultados de un ensayo multicéntrico se presentan en varias publicaciones (Huston 1996, Johansen 1999). A menudo las publicaciones redundantes no tienen referencias cruzadas entre sí (Bailey 2002, Barden 2003) y hay ejemplos donde dos artículos que reportan el mismo ensayo no comparten un único autor común (Gøtzsche 1989, Tramèr 1997). Por lo tanto, puede ser difícil o imposible para los revisores determinar si dos artículos constituyen publicaciones duplicadas de un estudio o dos estudios separados sin contactar a los autores, lo cual puede sesgar el metanálisis de estos datos.

10.2.2.2 Sesgo de ubicación

Las investigaciones indican que varios factores relacionados con la accesibilidad de los resultados de los estudios se asocian con los tamaños del efecto en los ensayos. Por ejemplo, en una serie de ensayos en el campo de la medicina complementaria y alternativa, Pittler y cols. examinaron la relación entre el resultado del ensayo, la calidad metodológica y el tamaño de muestra, y las características de las revistas donde se publicaron estos ensayos (Pittler 2000). Encontraron que era más probable que los ensayos publicados en revistas de bajo o ningún impacto reportaran resultados significativos, comparados con los publicados en las revistas médicas principales de gran impacto y que la calidad de los ensayos también se asoció con la revista de publicación. De manera similar, algunos estudios indican que es más probable que los ensayos publicados en revistas en idioma inglés muestren efectos significativos fuertes, en comparación con los publicados en revistas que no están en idioma inglés (Egger 1997b), aunque este hallazgo no ha sido consistente (Moher 2000, Jüni 2002, Pham 2005); ver Sección 10.2.2.4.

El término “sesgo de ubicación” también se utiliza para referirse a la accesibilidad de los estudios basada en la indexación variable en las bases de datos electrónicas. Según la pregunta clínica, la selección de las bases de datos donde buscar puede sesgar la estimación

del efecto en un metanálisis. Por ejemplo, un estudio encontró que los ensayos publicados en revistas que no estaban indexadas en MEDLINE pudieran mostrar efectos más beneficiosos que los ensayos publicados en revistas indexadas en MEDLINE (Egger 2003). Otros estudio de 61 metanálisis encontró que, en general, los ensayos publicados en revistas indexadas en EMBASE pero no en MEDLINE reportaron estimaciones del efecto más pequeñas que los indexados en MEDLINE, pero donde el riesgo de sesgo puede ser menor, debido a la menor prevalencia de ensayos únicos en EMBASE (Sampson 2003). Como se señaló anteriormente, estos hallazgos pueden variar de forma significativa según el tema que se examine.

Finalmente, una forma de sesgo de ubicación es el sesgo regional o de países desarrollados. Las investigaciones que apoyan las pruebas de este sesgo indican que es más probable que los estudios publicados en ciertos países produzcan investigaciones que muestren efectos significativos de las intervenciones. Vickers y cols. demostraron la posible existencia de este sesgo (Vicker 1998).

10.2.2.3 Sesgo de citación

El examen cuidadoso de las listas de referencias de los artículos se utiliza ampliamente para identificar artículos adicionales que pueden ser relevantes, aunque hay pocas pruebas que apoyan esta metodología. El problema con este enfoque es que el acto de citar trabajos previos está lejos del objetivo, por lo que la recuperación de la bibliografía mediante el examen de las listas de referencias puede producir una muestra sesgada de estudios. Existen muchas posibles motivaciones para citar un artículo. Brooks entrevistó a autores académicos de varias facultades de la Universidad de Iowa y les preguntó los motivos para citar cada referencia en uno de los artículos recientes de los autores (Brooks 1985). Ser persuasivos, es decir, el deseo de convencer a los revisores pares y fundamentar su propio punto de vista surgió como el motivo más importante para citar los artículos. Brooks concluyó que los autores recomiendan sus propias opiniones y utilizan la bibliografía para justificar su punto de vista: “Los autores se pueden definir como partidarios intelectuales de sus propias opiniones y por ello escudriñan la bibliografía para justificarse” (Brooks 1985).

En el análisis de Gøtzsche de los ensayos de fármacos anti-inflamatorios no esteroideos en la artritis reumatoide, fue más probable que se citaran los ensayos que demostraron un efecto superior del nuevo fármaco, comparados con los ensayos con resultados negativos (Gøtzsche 1987). Se mostraron resultados semejantes en un análisis de ensayos clínicos de enfermedades hepatobiliares (Kjaergard 2002). De forma similar, los ensayos de disminución del colesterol para prevenir la cardiopatía coronaria se citaron casi seis veces más si apoyaban la disminución del colesterol (Ravnskov 1992). También puede ocurrir la citación exagerada de estudios que no apoyan la intervención. Hutchinson y cols. examinaron revisiones sobre la efectividad de las vacunas pneumocócicas y encontraron que fue más probable que se citaran ensayos que no apoyaban la intervención, comparados con los que mostraban que la vacuna funcionaba (Hutchison 1995).

El sesgo de citación puede afectar la bibliografía “secundaria”. Por ejemplo, el *ACP Journal Club* tiene como objetivo resumir los artículos originales y de revisión de manera que los médicos puedan mantenerse al tanto de las últimas pruebas. Sin embargo, Carter y cols. encontraron que era más probable que se resumieran los ensayos con un resultado positivo, después de controlar por otros motivos para la selección (Carter 2006). Si es más probable que se resuman los resultados positivos, es más probable que los mismos se encuentren y, por lo tanto, es más probable que se incluyan en una revisión sistemática y sesguen así los hallazgos de la revisión.

10.2.2.4 Sesgo de idioma

A menudo las revisiones se han basado exclusivamente en estudios publicados en inglés. Por ejemplo, 36 metanálisis informados en revistas principales de medicina general en idioma inglés desde 1991 hasta 1993, 26 (72%) limitaron su búsqueda a los estudios informados en

inglés (Grégoire 1995). Esta tendencia puede estar en proceso de cambio. Una revisión reciente de 300 revisiones sistemáticas encontró que aproximadamente el 16% de las revisiones se limitaron a ensayos publicados en inglés; comparadas con las revisiones Cochrane, fue más probable que las revisiones sistemáticas publicadas en revistas de papel reportaran la limitación de sus búsquedas a ensayos publicados en inglés (Moher 2007). Además, de las revisiones con un enfoque terapéutico fue más probable que las revisiones Cochrane informaran que no tuvieron limitaciones de idioma, comparadas con las revisiones no Cochrane (62% vs. 26%) (Moher 2007).

Los investigadores que trabajan en países que no son de habla inglesa publicarán algunos de sus trabajos en revistas locales (Dickersin 1994). Es concebible que sea más probable que los autores publiquen en una revista internacional en inglés si los resultados son positivos, mientras que los hallazgos negativos se publican en una revista local. Lo anterior se demostró para la bibliografía en idioma alemán (Egger 1997b).

Por lo tanto, es posible introducir sesgo en revisiones basadas exclusivamente en informes en inglés (Grégoire 1995, Moher 1996). Sin embargo, las investigaciones que examinan este aspecto son contradictorias. En un estudio de 50 revisiones que utilizaron búsquedas exhaustivas en la bibliografía e incluyeron ensayos que estaban en inglés y que no estaban en inglés, Jüni y cols. informaron que fue más probable que los ensayos que no estaban en inglés produjeran resultados significativos al nivel de $p < 0,05$, mientras que las estimaciones de los efectos de la intervención fueron, como promedio 16% (IC del 95%: 3% al 26%) más beneficiosos en los ensayos que no estaban en inglés que en los que se publicaron en este idioma (Jüni 2002). Por el contrario, Moher y cols. examinaron el efecto de la inclusión o exclusión de los ensayos en inglés en dos metanálisis y en general encontraron que la exclusión de los ensayos reportados en un idioma diferente del inglés no afectó significativamente los resultados de los metanálisis (Moher 2003). Estos resultados fueron similares cuando el análisis se limitó a los metanálisis de ensayos de fármacos convencionales. Sin embargo, cuando el análisis se realizó por separado para los metanálisis de ensayos de medicina alternativa y complementaria, el tamaño del efecto de los metanálisis disminuyó significativamente al excluir los informes en idiomas diferentes del inglés (Moher 2003).

El grado y los efectos del sesgo de idioma pueden haber disminuido recientemente debido al cambio hacia la publicación de estudios en inglés. En 2006 Galandi y cols. informaron de una gran disminución en el número de ensayos aleatorios publicados en revistas de atención sanitaria en idioma alemán, con menos de dos ensayos aleatorios publicados por revista y por año después de 1999 (Galandi 2006). Aunque la repercusión potencial de los estudios publicados en idiomas diferentes del inglés en un metanálisis puede ser mínima, es difícil predecir en qué casos esta exclusión puede sesgar una revisión sistemática. Los revisores pueden decidir realizar las búsquedas sin limitaciones de idioma y las decisiones acerca de la inclusión de informes de idiomas diferentes del inglés se deben tomar individualmente caso a caso.

10.2.2.5 Sesgo de informe de resultado

En muchos estudios se registran varias medidas de resultado pero no se reportan todas (Pocock 1987, Tannok 1996). La selección de las medidas de resultado que se informan puede estar influenciada por los propios resultados, lo que puede hacer que los que se publiquen sean erróneos. Por ejemplo, dos análisis separados (Mandel 1987, Cantekin 1991) de un ensayo controlado a doble ciego con placebo que evaluó la eficacia de amoxicilina en niños con otitis media no supurativa establecieron conclusiones opuestas debido a que se dieron diferentes “ponderaciones” a varias medidas de resultado que se evaluaron en el estudio. Este desacuerdo se discutió de forma pública, ya que se acompañó de acusaciones de procedimientos incorrectos contra el equipo que produjo los hallazgos favorables a la amoxicilina. El director de este equipo había recibido un importante apoyo fiscal en becas de investigación y como honorarios personales de los fabricantes de amoxicilina (Rennie 1991).

El anterior es un buen ejemplo de cómo la confianza en los datos que los investigadores deciden presentar puede dar lugar a distorsiones (Anonymous 1991). Dicho "sesgo de informe de resultado" puede ser particularmente importante para los efectos adversos. Hemminki examinó los informes de ensayos clínicos enviados por las compañías farmacéuticas a las autoridades que expiden licencias en Finlandia y Suecia y encontró que los ensayos no publicados proporcionaron información sobre los efectos adversos más a menudo que los ensayos publicados (Hemminki 1980). Desde entonces muchos otros estudios han mostrado que el informe de los eventos adversos y los resultados de seguridad en los ensayos clínicos a menudo es inadecuado y selectivo (Ioannidis 2001, Melander 2003, Heres 2006). Un grupo de Canadá, Dinamarca y el RU fue recientemente el primero en realizar investigaciones empíricas acerca del informe selectivo de los resultados del estudio (Chan 2004a, Chan 2004b, Chan 2005). Estos estudios se describen en el Capítulo 8 (Sección 8.14), junto con una discusión más detallada del sesgo de informe de resultado.

10.3 Cómo evitar los sesgos de informe

10.3.1 Implicaciones de las pruebas relacionadas con los sesgos de informe

Las pruebas convincentes de la presencia de sesgos de informe, descritas en la Sección 10.2, demuestran la necesidad de buscar de forma exhaustiva los estudios que cumplan con los criterios de elegibilidad para una revisión Cochrane. Los revisores deben asegurar que se buscó en múltiples fuentes, por ejemplo una búsqueda solamente en MEDLINE no se consideraría suficiente. Las fuentes de información y los métodos de búsqueda se describen en detalle en el Capítulo 6. No obstante, las búsquedas exhaustivas no eliminan necesariamente el sesgo. Los revisores deben tener en mente, por ejemplo, que los informes de los estudios pueden presentar selectivamente los resultados, que las listas de referencias pueden citar las fuentes de forma selectiva y que puede ser difícil detectar las publicaciones duplicadas de los resultados. Además, la disponibilidad de la información del estudio puede estar sujeta al sesgo de lapso de tiempo, particularmente en áreas de investigación donde los cambios ocurren de forma rápida. A continuación se discuten dos maneras adicionales de reducir o posiblemente evitar los sesgos de informe: la inclusión de estudios no publicados y el uso de registros de ensayos.

10.3.2 Inclusión de estudios no publicados en las revisiones sistemáticas

Está claro que el sesgo de publicación es una amenaza importante a la validez de cualquier tipo de revisión, pero particularmente en el caso de las revisiones no sistemáticas narrativas. La obtención e inclusión de datos de ensayos no publicados parece ser una manera obvia de evitar este problema. Hopewell y cols. realizaron una revisión de estudios que comparaban el efecto de la inclusión o exclusión de la literatura "gris" (definida aquí como los informes producidos por todos los niveles gubernamentales, académicos, de negocios e industriales en formato impreso y electrónico, pero que no están controlados por editoriales comerciales) en los metanálisis de ensayos aleatorios (Hopewell 2007b). Se incluyeron cinco estudios (Fergusson 2000, McAuley 2000, Burdett 2003, Hopewell 2004), que en general mostraron que los ensayos publicados observaron un mayor efecto de la intervención que los ensayos de la literatura gris. Un metanálisis de tres de estos estudios indicó que, como promedio, los ensayos publicados mostraron un efecto de la intervención 9% mayor que los de la literatura gris (Hopewell 2007b).

La inclusión de los datos de estudios no publicados puede introducir un sesgo por sí misma. Los estudios localizados pueden ser una muestra no representativa de todos los estudios no publicados. Los estudios no publicados pueden tener una calidad metodológica menor que los estudios publicados. Un estudio de 60 metanálisis que incluyó ensayos publicados y no publicados encontró que fue menos probable que los ensayos no publicados ocultaran adecuadamente la asignación de la intervención y cegaran las evaluaciones de resultados

(Egger 2003). Por el contrario, Hopewell y cols. no encontraron diferencias en la calidad del reporte de esta información (Hopewell 2004).

Un problema adicional se relaciona con el consentimiento de proporcionar los datos de los investigadores de los estudios no publicados localizados. Este consentimiento puede depender de los hallazgos del estudio, donde los resultados más favorables se proporcionan con más rapidez. Nuevamente, lo anterior puede sesgar los hallazgos de una revisión sistemática. Es interesante señalar que, cuando Hetherington y cols., en un gran esfuerzo por obtener información acerca de ensayos no publicados en medicina perinatal, abordaron a 42 000 obstetras y pediatras de 18 países, identificaron solamente 18 ensayos no publicados que habían finalizado hacía mas de dos años (Hetherington 1989).

Se envió un cuestionario que evaluaba las actitudes hacia la inclusión de datos no publicados a los autores de 150 metanálisis y a los editores de las revistas que los publicaron (Cook 1993). Los investigadores y los editores difirieron en sus puntos de vista acerca de incluir datos no publicados en los metanálisis. Fue evidente el apoyo del uso de material no publicado entre una clara mayoría de los autores de metanálisis (78%), mientras que los editores de revistas se mostraron menos convencidos (47%) (Cook 1993). Recientemente se repitió este estudio, centrado en la inclusión de la literatura gris en las revisiones sistemáticas, y se encontró un aumento en la aceptación de la inclusión de la literatura gris. Aunque se mantuvieron las diferencias entre los grupos (autores de revisiones sistemáticas: 86%, editores: 69%), las mismas disminuyeron en comparación con los datos presentados por Cook y cols. (Tetzlaff 2006).

Entre los motivos para rechazar la inclusión de la literatura gris estaba la falta de revisión por pares de la bibliografía no publicada. Sin embargo, se debe tener en cuenta que el proceso de arbitraje no siempre ha sido una manera exitosa de asegurar que los resultados publicados sean válidos (Godlee 1999). El equipo involucrado en la preparación de una revisión Cochrane debe tener al menos un nivel de experiencia similar al de un revisor par de una revista para evaluar los estudios no publicados. Por otra parte, los metanálisis de datos no publicados procedentes de fuentes interesadas son claramente un motivo de inquietud.

10.3.3 Registros de ensayos y sesgo de publicación

En septiembre de 2004 varias revistas médicas importantes pertenecientes al International Committee of Medical Journal Editors (ICMJE) anunciaron que no publicarían más ensayos que no estuvieran registrados desde su inicio (Abbasi 2004). Todos los ensayos que comenzaron el reclutamiento de los participantes después de septiembre de 2005 debían estar registrados en un registro público de ensayos en el momento o antes del inicio del reclutamiento para ser considerados para su publicación en estas revistas. El ICMJE describió los registros “aceptables”: posibilidad de búsqueda electrónica, con libre acceso para el público, abiertos a todas las personas que registraran, y administrados por una organización sin ánimo de lucro. De manera similar, el ICMJE le pide a los autores de ensayos clínicos que cumplan con un conjunto mínimo de datos propuestos por la Organización Mundial de la Salud.

Si esta iniciativa que se debió haber hecho mucho tiempo antes es exitosa, tiene la posibilidad de reducir significativamente los efectos del sesgo de publicación. Sin embargo, lo anterior dependería de que los revisores identifiquen todos los ensayos relevantes mediante búsquedas en registros de ensayos en línea, y también de que los resultados de los ensayos no publicados, identificados mediante los registros estén disponibles para ellos. Las iniciativas para obligar al registro de los resultados de los ensayos están actualmente en sus primeros estadios, evolucionan rápidamente y deben tener una repercusión sobre la accesibilidad de estos datos. Aunque han surgido pruebas que indican que algunos de los campos de datos solicitados en los registros están incompletos (Zarin 2005), es probable que esta situación mejore con el tiempo. En este momento no está claro el grado en el cual el registro de los ensayos facilitará el trabajo de los revisores Cochrane. Para consultar sobre la búsqueda en los registros de ensayos, ver Capítulo 6 (Sección 6.2.3).

10.4 Detección de los sesgos de informe

10.4.1 Gráficos de embudo

Un gráfico de embudo es un gráfico sencillo que ilustra acerca de la dispersión de las estimaciones de los efectos de la intervención en estudios individuales contra alguna medida del tamaño o la precisión de cada estudio. En común con los gráficos forestales, es más frecuente colocar las estimaciones del efecto en la escala horizontal y por lo tanto, la medición del tamaño del estudio sobre el eje vertical. Esta forma es la opuesta de la presentación gráfica convencional de los gráficos de dispersión, en los cuales el resultado (p.ej. el efecto de la intervención) se coloca en el eje vertical y la co-variable (p.ej. el tamaño del estudio) en el eje horizontal.

El nombre “gráfico de embudo” proviene del hecho de que la precisión de la estimación del efecto de la intervención aumenta cuando se incrementa el tamaño del estudio. Por lo tanto, las estimaciones del efecto de los estudios pequeños se dispersarán más ampliamente en la parte inferior del gráfico, y la dispersión se estrechará entre los estudios más grandes. En ausencia de sesgo el gráfico se debe parecer aproximadamente a un embudo simétrico (invertido). Lo anterior se ilustra en el Panel A de la Figura 10.4.a, en el cual las estimaciones del efecto en los estudios más grandes están cerca del verdadero odds ratio de la intervención de 0,4.

Si existe sesgo, por ejemplo debido a estudios más pequeños sin efectos estadísticamente significativos (mostrados como círculos abiertos en la Figura 10.4.a, Panel A) que no se han publicado, dará lugar a una apariencia asimétrica del gráfico en embudo con una brecha en una esquina inferior del gráfico (Panel B). En esta situación el efecto calculado en un metanálisis tenderá a sobrestimar el efecto de la intervención (Egger 1997a, Villar 1997). Mientras más pronunciada sea la asimetría, será más probable que la cantidad de sesgo sea significativa.

Los gráficos de embudo se utilizaron por primera vez en investigaciones educativas y en psicología, donde las estimaciones de los efectos se colocaron contra el tamaño total de la muestra (Light 1984). Actualmente se recomienda de forma habitual que se muestre el error estándar de la estimación del efecto de la intervención, en lugar del tamaño total de la muestra, en el eje vertical (Sterne 2001). Lo anterior se debe a que el poder estadístico de un ensayo está determinado por otros factores además del tamaño de la muestra, como el número de participantes que presentan el evento para los resultados dicotómicos y la desviación estándar de las respuestas para los resultados continuos. Por ejemplo, es menos probable que un estudio con 100.000 participantes y 10 eventos muestre un efecto de la intervención que sea estadísticamente significativo que un estudio con 1000 participantes y 100 eventos. El error estándar resume estos otros factores. Al incluir los errores estándar en una escala inversa, los estudios mayores o con mayor poder estadístico se colocan hacia la parte superior del gráfico. Otra ventaja potencial del uso de los errores estándar es que es posible mostrar una región triangular sencilla dentro de la cual se espera que esté el 95% de los estudios si no hay sesgo ni heterogeneidad. Estas regiones se incluyen en la Figura 10.4.a. Los gráficos de embudo de las estimaciones del efecto contra sus errores estándar (en una escala inversa) se pueden crear mediante RevMan. Es posible incluir en el gráfico una región triangular de confianza del 95% según un metanálisis de efectos fijos y diferentes símbolos en el gráfico permiten identificar los estudios en diferentes subgrupos.

El sesgo de publicación no provoca necesariamente asimetría en los gráficos de embudo. En ausencia de cualquier efecto de la intervención, la publicación selectiva basada solamente en el valor de la P dará lugar a un gráfico de embudo simétrico en el cual es más probable que se publiquen los estudios que se encuentran en el lado extremo izquierdo o derecho, comparados con los que están en el centro. Ello podría sesgar la estimación de la varianza en la heterogeneidad entre estudios.

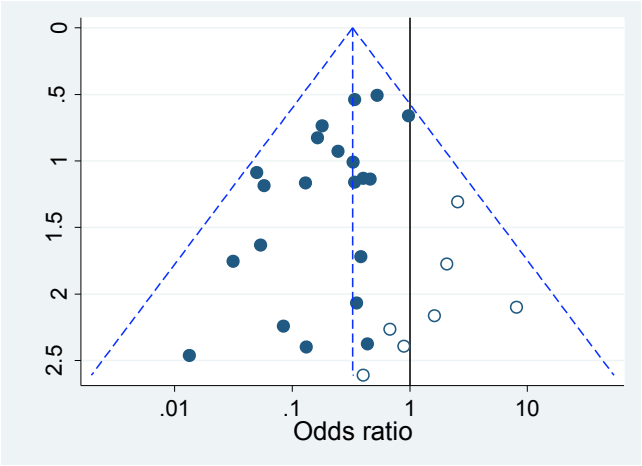
Las medidas de proporción de los efectos de la intervención (como los odds ratios y los cocientes de riesgos) se deben colocar en una escala logarítmica, lo que asegura que los efectos de la misma magnitud pero con dirección opuesta (por ejemplo, odds ratios de 0,5 y 2) estén equidistantes de 1,0. Para los resultados medidos en una escala continua (numérica) (p.ej. presión arterial, puntuación de depresión) los efectos de la intervención se miden como diferencias de medias o diferencias de medias estandarizadas, las cuales por lo tanto se deben utilizar en el eje horizontal de los gráficos de embudo. Hasta donde se conoce, ninguna investigación empírica ha examinado la elección de los ejes para los gráficos de embudo con resultados continuos. Para las diferencias de medias el error estándar es aproximadamente proporcional al inverso de la raíz cuadrada del número de participantes, por lo que parece ser una decisión indiscutible para el eje vertical.

Algunos autores han argumentado que la interpretación visual de los gráficos de embudo es demasiado subjetiva para ser útil. En particular, Terrin y cols. encontraron que los investigadores sólo tienen una capacidad limitada para identificar correctamente los gráficos de embudo de los metanálisis con riesgo de sesgo de publicación (Terrin 2005).

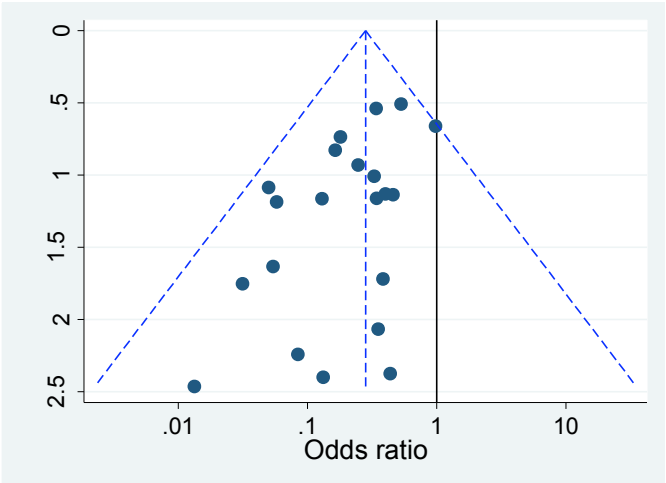
Un problema adicional importante que tienen los gráficos de embudo es que algunas estimaciones del efecto (p.ej. odds ratios y diferencias de medias estandarizadas) se correlacionan de forma natural con sus errores estándar y pueden producir una asimetría espuria en un gráfico de embudo. Este problema se comenta con más detalle en la Sección 10.4.3.

Figura 10.4.a: Gráficos de embudo hipotéticos
Panel A: gráfico simétrico en ausencia de sesgo. Panel B: gráfico asimétrico en presencia de sesgo de informe. Panel C: gráfico asimétrico en presencia de sesgo debido a que algunos estudios más pequeños (círculos abiertos) son de menor calidad metodológica, por lo que producen estimaciones exageradas del efecto de la intervención.

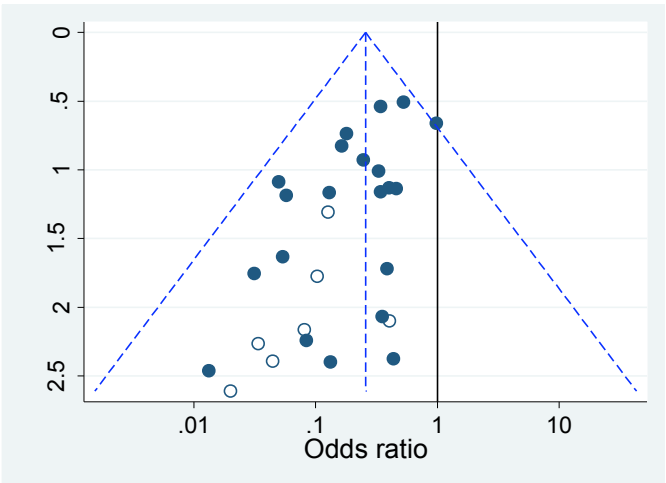
Panel A



Panel B



Panel C



10.4.2 Diferentes motivos para la asimetría del gráfico de embudo

Aunque la asimetría del gráfico de embudo se ha igualado hace tiempo al sesgo de publicación (Light 1984, Begg 1988), el gráfico de embudo se debe considerar un medio genérico de mostrar los *efectos de los estudios pequeños*, que significa que la estimación de los efectos de la intervención en los estudios más pequeños tiende a diferir de las estimaciones de los estudios mayores (Sterne 2000). Los efectos de los estudios pequeños se pueden deber a motivos diferentes del sesgo de publicación (Egger 1997a, Sterne 2000). Algunos de éstos se muestran en la Tabla 10.4.a.

Las diferencias en la calidad metodológica son una fuente potencial importante de asimetría del gráfico de embudo. Los estudios más pequeños tienden a realizarse y analizarse con menor rigor metodológico que los estudios más grandes (Egger 2003).Los ensayos de menor calidad tienden a mostrar mayores efectos de la intervención (Schulz 1995). Por lo tanto, los ensayos que hubieran sido “negativos”, si se realizaran y analizaran adecuadamente, pueden convertirse en “positivos” (Figura 10.4.a, Panel C).

La heterogeneidad verdadera en los efectos de la intervención también puede dar lugar a la asimetría del gráfico de embudo. Por ejemplo, se puede observar un beneficio significativo solamente en los pacientes con alto riesgo para el resultado que está afectado por la intervención y habitualmente es más probable que estos pacientes con alto riesgo se incluyan en los estudios pequeños iniciales (Davey Smith 1994, Glasziou 1995). Además, los ensayos pequeños se realizan generalmente antes de que se desarrollen los más grandes y en los años del ensayo es posible que el tratamiento estándar haya mejorado (lo que da lugar a menores efectos de la intervención en los ensayos más grandes). Adicionalmente, algunas intervenciones pueden haberse implementado con menos cuidado en los ensayos más grandes, por lo que pueden provocar estimaciones más pequeñas del efecto de la intervención (Stuck 1998). Finalmente, es posible que un gráfico de embudo aparezca solamente por azar. Terrin y cols. han indicado que el gráfico de embudo no es apropiado para los metanálisis heterogéneos y han llamado la atención sobre la premisa de que los estudios provienen de una población subyacente única definida por los generadores del gráfico de embudo (Light 1984, Terrin 2003).

Una mejoría propuesta (Peters 2008) para los gráficos de embudo es incluir líneas de contorno que correspondan a “valores clásicos” percibidos como de significación estadística (p = 0,01; 0,05; 0,1 etc.). Lo anterior permite destacar la significación estadística de las estimaciones de los estudios, así como las áreas de estudios faltantes. Estos gráficos de embudo de “contorno mejorado” pueden ayudar a los revisores a diferenciar la asimetría debida a sesgo de publicación de la debida a otros factores. Por ejemplo, si parecer que hay estudios que faltan en áreas de no significación estadística (ver Figura 10.4.b, Panel A como ejemplo) entonces se incrementa la posibilidad de que la asimetría se deba al sesgo de publicación. Por el contrario, si los estudios que supuestamente faltan están en áreas de mayor significación estadística (ver Figura 10.4.b, Panel B como ejemplo), indicaría que es más probable que la asimetría se deba a factores diferentes del sesgo de publicación (ver Tabla 10.4.a). Si no hay estudios estadísticamente significativos entonces es posible que el sesgo de publicación no sea una explicación plausible para la asimetría del gráfico de embudo (Ioannidis 2007b).

Al interpretar los gráficos de embudo los autores de revisiones sistemáticas deben distinguir los posibles motivos para la asimetría del gráfico de embudo enumerados en la Tabla 10.4.a. El conocimiento de la intervención particular y las circunstancias en las cuales se implementó la misma en los diferentes estudios pueden ayudar a identificar la verdadera heterogeneidad como causa de la asimetría del gráfico de embudo. Aún existe la inquietud de que la interpretación visual de los gráficos de embudo es inherentemente subjetiva. Por lo tanto, a continuación se discuten las pruebas estadísticas para la asimetría del gráfico de embudo y el grado en el cual pueden ayudar a la interpretación objetiva de los mismos. Cuando los revisores estén preocupados porque los efectos de los estudios pequeños puedan influir sobre

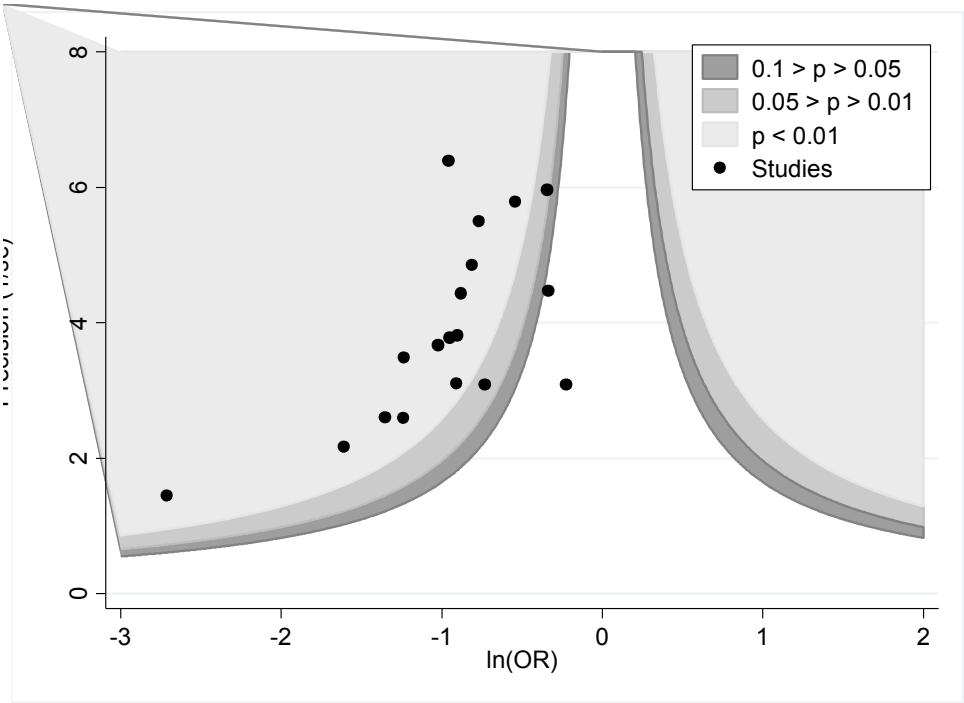
los resultados de un metanálisis, podría ser interesante realizar análisis de sensibilidad para explorar la solidez de las conclusiones del metanálisis para diferentes suposiciones sobre de las causas la asimetría del gráfico de embudo, las cuales se discuten en la Sección 10.4.4.

Tabla 10.4.a: Posibles fuentes de asimetría en los gráficos de embudo
Adaptado de Egger y cols. (Egger 1997a).

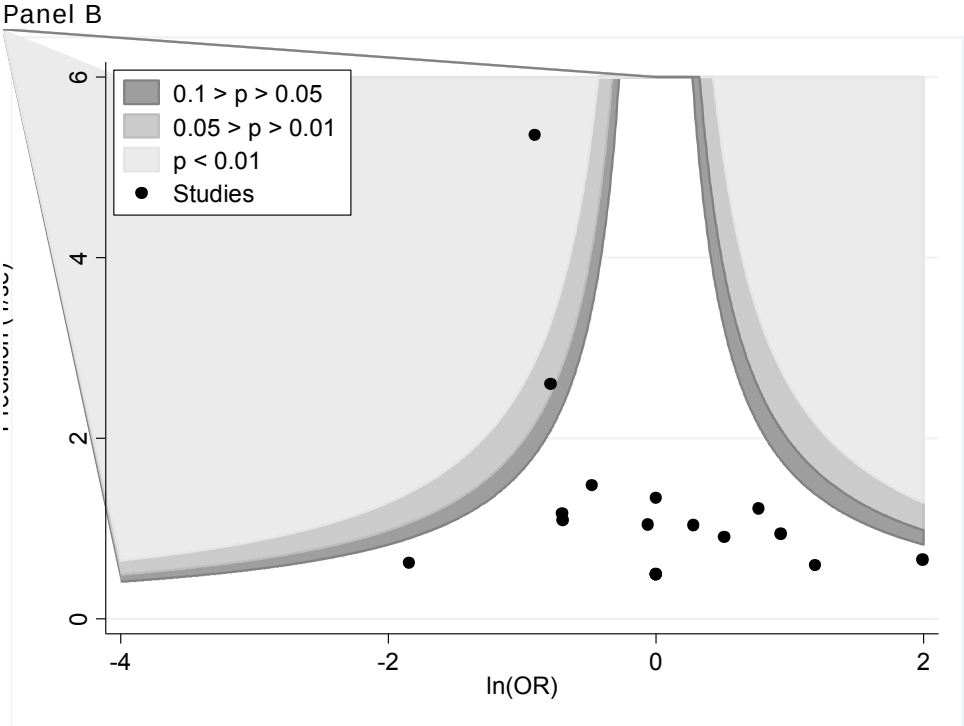
1. Sesgos de selección: Sesgo de publicación: Sesgo de publicación demorada (también conocido como “lapso de tiempo” o “de trámite”). Sesgo de ubicación: Sesgo de idioma; Sesgo de citación; Sesgo de publicación múltiple. Informe selectivo de resultado.
2. Calidad metodológica deficiente que da lugar a efectos espurio exagerados en los estudios más pequeños: Diseño metodológico deficiente; Análisis inadecuado; Fraude.
3. Heterogeneidad verdadera: El tamaño del efecto difiere según el tamaño del estudio (por ejemplo, debido a diferencias en la intensidad de las intervenciones o a diferencias en el riesgo subyacente entre los estudios de tamaños diferentes).
4. Por artefacto: En algunas circunstancias (ver Sección 10.4.3), la variación en la muestra puede dar lugar a una asociación entre el efecto de una intervención y su error estándar.
5. Por azar.

Figura 10.4.b: Gráficos de embudo de contorno mejorado
Panel A: existe una indicación de estudios faltantes hacia el lado derecho del gráfico, mayormente en el área de no significación (es decir, el área blanca donde p > 0,1) donde el sesgo de publicación es una explicación posible. Panel B: existe una indicación de estudios que faltan en el lado inferior izquierdo del gráfico. Como la mayor parte de este área contiene las regiones de alta significación estadística (ello es, indicada por el sombreado oscuro), se reduce la posibilidad de que el sesgo de publicación sea la causa subyacente de esta asimetría del embudo.

Panel A



Panel B



10.4.3 Pruebas para la asimetría del gráfico de embudo

Una prueba para la asimetría del gráfico de embudo (efectos de los estudios pequeños) examina formalmente si la asociación entre la estimación de los efectos de la intervención y una medida del tamaño del estudio (como el error estándar del efecto de la intervención) es mayor que lo que se podría esperar que ocurra por azar. Para los resultados medidos en una escala continua (numérica) este planteamiento es relativamente sencillo. Mediante un enfoque propuesto por Egger y cols. (Egger 1997a), es posible realizar una regresión lineal de las estimaciones del efecto de la intervención con sus errores estándar, con una ponderación de 1/(varianza de la estimación del efecto de la intervención). El procedimiento busca una relación lineal directa entre el efecto de la intervención y su error estándar. Bajo la hipótesis nula de ningún efecto de los estudios pequeños (p.ej. Panel A en la Figura 10.4.a) dicha línea debería ser vertical. A mayor asociación entre el efecto de la intervención y el error estándar (p.ej. como el Panel B en la Figura 10.4.a), más se alejará la pendiente de la vertical. Hay que señalar que la ponderación es importante para asegurar que las estimaciones de la regresión no estén dominadas por los estudios más pequeños.

Cuando los resultados son dicotómicos y los efectos de la intervención se expresan como odds ratios, el enfoque propuesto por Egger y cols. (Egger 1997a) corresponde a una regresión lineal del logaritmo del odds ratio con su error estándar, ponderado por la inversa de la varianza del logaritmo del odds ratio (Sterne 2000). Éste ha sido el enfoque más ampliamente utilizado y citado para probar la asimetría del gráfico de embudo. Desafortunadamente hay problemas estadísticos con el mismo, debido a que el error estándar del logaritmo del odds ratio está vinculado matemáticamente al tamaño del odds ratio, incluso en ausencia de efectos de los estudios pequeños (Irwig 1998) (ver Deeks y cols. para una explicación algebraica de este fenómeno [Deeks 2005]). Lo anterior puede provocar que los gráficos de embudo donde se utilice el logaritmo de los odds ratios (o los odds ratios en una escala logarítmica) parezcan asimétricos y puede significar que los valores de P provenientes de la prueba de Egger y cols. son demasiado pequeños, lo que da lugar a resultados falsos positivos de la prueba. Estos problemas son especialmente proclives a ocurrir cuando la intervención tiene un gran efecto, existe una heterogeneidad importante entre los estudios, hay pocos eventos por estudio, o cuando los estudios son de tamaño similar.

Por lo tanto, varios autores han propuesto pruebas alternativas para la asimetría del gráfico de embudo, las cuales se resumen en la Tabla 10.4.b. Debido a que es imposible conocer el mecanismo exacto del sesgo de publicación, se requieren estudios de simulación (en los cuales las pruebas se evalúan en un gran número de conjuntos de datos generados por ordenador) para evaluar las características de las pruebas en un rango de suposiciones sobre del mecanismo del sesgo de publicación (Sterne 2000, Macaskill 2001, Harbord 2006, Peters 2006, Schwarzer 2007). El estudio más exhaustivo (en términos de escenarios examinados, simulaciones realizadas y el rango de pruebas realizadas) lo informó Rücker et al. (Rücker 2008). Los resultados de éste y otros estudios de simulaciones publicados reportan las recomendaciones señaladas más adelante sobre las pruebas de la asimetría del gráfico de embudo. Aunque los estudios de simulación proporcionan apreciaciones útiles, evalúan inevitablemente circunstancias que difieren de un metanálisis determinado, por lo que sus resultados se deben interpretar con cuidado.

La mayor parte de este trabajo metodológico se ha centrado en los efectos de la intervención medidos como odds ratios. Aunque parece plausible esperar que surjan problemas similares para los efectos de la intervención medidos como cocientes de riesgos o diferencias de medias estandarizadas, se requieren investigaciones adicionales de estas situaciones.

Actualmente se mantiene el debate sobre la representatividad de los valores paramétricos utilizados en los estudios de simulación y los mecanismos utilizados para simular el sesgo de publicación y los efectos de los estudios pequeños, que a menudo se seleccionan con muy poca justificación explícita. Aún no se han examinado algunas variaciones potencialmente útiles en las diferentes pruebas. Por lo tanto, no es posible hacer recomendaciones definitivas

sobre la selección de las pruebas para la asimetría del gráfico de embudo. Sin embargo, es posible identificar tres pruebas que deben considerar los revisores para analizar la asimetría del gráfico de embudo.

Ninguna de las pruebas descritas aquí está implementada en RevMan. Es recomendable consultar con un estadístico para su implementación.

Tabla 10.4.b: Pruebas propuestas para analizar la asimetría del gráfico de embudo
N_{tot} es el tamaño de muestra total, N_E y N_C son los tamaños de muestra de los grupos de intervención experimental y control, S es el número total de eventos entre ambos grupos y F = N_{tot} – S. Note que sólo las tres primeras de estas pruebas (Begg 1994, Egger 1997a, Tang 2000) se pueden utilizar para los resultados continuos.

Referencia	Bases de la prueba
(Begg 1994)	Correlación de rangos entre el efecto de la intervención estandarizado y su error estándar.
(Egger 1997a)	Regresión lineal de la estimación del efecto de la intervención contra su error estándar, ponderado por la inversa de la varianza de la estimación del efecto de la intervención.
(Tang 2000)	Regresión lineal de la estimación del efecto de la intervención sobre $1/\sqrt{N_{tot}}$, con ponderaciones N_{tot}.
(Macaskill 2001)*	Regresión lineal de la estimación del efecto de la intervención sobre N_{tot}, con ponderaciones $S \times F/N_{tot}$.
(Deeks 2005)*	Regresión lineal del logaritmo del odds ratio sobre $1/\sqrt{ESS}$ con ponderaciones ESS, donde el tamaño de muestra efectivo es ESS = $4N_E \times N_C / N_{tot}$.
(Harbord 2006)*	Versión modificada de la prueba propuesta por Egger y cols., basada en la “puntuación” (O - E) y la “varianza de la puntuación” (V) del logaritmo del odds ratio.
(Peters 2006)*	Regresión lineal de la estimación del efecto de la intervención sobre $1/N_{tot}$, con ponderaciones $S \times F/N_{tot}$.
(Schwarzer 2007)*	Prueba de correlación de rangos que utiliza la media y la varianza de la distribución hipergeométrica no central.
(Rücker 2008)	Prueba basada en la transformación del seno del arco de los riesgos observados, con modelado explícito de la heterogeneidad entre los estudios.

* Prueba formulada en términos de los odds ratios, pero puede ser aplicable a otras medidas de efecto de la intervención.

10.4.3.1 Recomendaciones para las pruebas de asimetría del gráfico de embudo para todo tipo de resultado:

- Como una regla empírica, las pruebas para la asimetría del gráfico de embudo se deben utilizar solamente cuando se hayan incluido al menos 10 estudios en el metanálisis, ya que cuando hay menos estudios el poder estadístico de la prueba es demasiado bajo para distinguir el azar de la verdadera asimetría.
- Las pruebas para la asimetría del gráfico de embudo no se deben utilizar si todos los estudios son de tamaño similar (similares errores estándar de las estimaciones del efecto de la intervención). Sin embargo, no se conocen pruebas de estudios de simulación que proporcionen guías específicas sobre cuándo los tamaños de los estudios se deben considerar “demasiado similares”.
- Para interpretar los resultados de las pruebas para la asimetría del gráfico de embudo se debe realizar la inspección visual del mismo. Por ejemplo, ¿los estudios más pequeños dan lugar a estimaciones del efecto de la intervención más o menos beneficiosas? ¿Existen estudios con estimaciones del efecto de la intervención marcadamente diferentes

(atípicos) o estudios que tienen una gran influencia en el metanálisis? ¿El valor pequeño de la P está causado por un solo estudio? El examen de un gráfico de embudo de contorno mejorado, como se esboza en la Sección 10.4.1, puede ayudar adicionalmente a la interpretación del resultado de una prueba.

- Cuando hay pruebas de efectos de los estudios pequeños, el sesgo de publicación sólo se debe considerar como una de varias posibles explicaciones (ver Tabla 10.4.a). Aunque los gráficos de embudo y las pruebas de asimetría del gráfico de embudo pueden alertar a los revisores sobre un problema que necesita considerarse, no proporcionan una solución a este problema.
- Finalmente, los revisores deben recordar que debido a que generalmente las pruebas tienen poco poder estadístico, incluso cuando una prueba no proporciona evidencias de asimetría del gráfico de embudo, no es posible excluir el sesgo (incluido el de publicación).

Para resultados continuos con efectos de la intervención medidos como diferencia de medias:

- La prueba propuesta por Egger (Egger 1997a) se puede utilizar para probar la asimetría del gráfico en embudo. Actualmente no hay motivos para preferir una de las pruebas propuestas más recientemente en esta situación, aunque sus ventajas y desventajas relativas no se han examinado formalmente. Aunque no se conocen investigaciones específicas sobre el poder estadístico de este enfoque en el caso de los resultados continuos, las consideraciones generales indican que el poder estadístico será mayor que para los resultados dicotómicos pero que no es recomendable utilizar el método con mucho menos de 10 estudios.

Para resultados dicotómicos con efectos de la intervención medidos como odds ratios:

- Las pruebas propuestas por Harbord y cols. (Harbord 2006) y Peters y cols. (Peters 2006) evitan la asociación matemática entre los odds ratios y sus errores estándar (y por lo tanto los resultados falsos positivos de las pruebas) que aparecen con la prueba de Egger y cols. cuando hay un efecto significativo de la intervención, a la vez que mantienen el poder estadístico, comparadas con las pruebas alternativas. Sin embargo, aún pueden aparecer resultados falsos positivos en presencia de heterogeneidad significativa entre los estudios.
- La prueba propuesta por Rücker y cols. (Rücker 2008) evita los resultados falsos positivos cuando hay un efecto significativo de la intervención y en presencia de heterogeneidad significativa entre los estudios. Como una regla empírica, cuando la varianza de la estimación de la heterogeneidad entre estudios del logaritmo de los odds ratios, tau cuadrado, es más de 0,1, sólo la versión de la prueba del seno del arco que incluye los efectos aleatorios (referida como “AS+RE” por Rücker y cols.) ha mostrado que funciona razonablemente bien. Sin embargo, la misma es ligeramente conservadora en ausencia de heterogeneidad y su interpretación es menos familiar debido a que se basa en una transformación del seno del arco. (Es de señalar que aunque esta recomendación se basa en la magnitud de tau cuadrado, otros factores, incluidos los tamaños de los diferentes estudios y su distribución, influyen en el rendimiento de la prueba. Actualmente no es posible incorporar dichos factores en estas recomendaciones).
- Cuando la varianza de la heterogeneidad, tau cuadrado, es menor de 0,1 se puede utilizar una de las pruebas propuestas por Harbord 2006, Peters 2006 o Rücker 2008. (En general el rendimiento de la prueba se deteriora cuando tau cuadrado aumenta).
- Siempre que sea posible los revisores deben especificar su estrategia de pruebas por adelantado (y señalar que la selección de las pruebas puede depender del grado de heterogeneidad observado). Se debe aplicar solamente una prueba, apropiada al contexto del metanálisis particular, de la lista de recomendadas anteriormente e informar solamente el resultado de la prueba seleccionada. La aplicación de dos o más pruebas no es recomendable porque los valores de P más extremos (los más grandes o los más pequeños) de un conjunto de pruebas no tienen una interpretación bien caracterizada.

Para resultados dicotómicos con efectos de la intervención medidos como cocientes de riesgo o diferencias de riesgo, y resultados continuos con efectos de la intervención medidos como diferencias de medias estandarizadas:

- Los problemas potenciales de los gráficos en embudo se han estudiado con menos profundidad para estas medidas de efecto que para los odds ratios, por lo que aún no hay guías sólidas disponibles.
- Generalmente los metanálisis de las diferencias de riesgos se consideran menos apropiados que los metanálisis que utilizan un cociente como medida de efecto (ver Capítulo 9, Sección 9.4.4.4). Por motivos similares, pocas veces son de interés los gráficos de embudo que utilizan las diferencias de riesgo. Si el cociente de riesgo (o el odds ratio) es constante entre los estudios, entonces un gráfico de embudo que utilice las diferencias de riesgo será asimétrico si los estudios más pequeños tienen un riesgo inicial mayor (o menor).

Según una encuesta de metanálisis publicados en la *Base de Datos Cochrane de Revisiones Sistemáticas*, estos criterios significan que las pruebas para la asimetría del gráfico de embudo se pueden utilizar solamente en una minoría de los metanálisis (Ioannidis 2007b).

Pruebas para las cuales no hay evidencia suficiente para recomendar su uso
Los siguientes comentarios se aplican a todas las medidas de intervención. La prueba propuesta por Begg y Mazumdar (Begg 1994) tiene los mismos problemas estadísticos, pero un poder estadístico menor, que la prueba de Egger y cols., por lo que no se recomienda. La prueba propuesta por Tang y Liu (Tang 2000) no se ha evaluado en estudios de simulación, mientras que la propuesta por Macaskill y cols. (Macaskill 2001) tiene un poder estadístico menor que las alternativas propuestas anteriormente. La prueba propuesta por Schwarzer y cols. (Schwarzer 2007) evita la asociación matemática entre los odds ratios y su error estándar, pero tiene un poder estadístico menor con respecto a las pruebas discutidas antes.

En el contexto de los metanálisis de los estudios de intervención considerados en este capítulo, es posible que la prueba propuesta por Deeks y cols. (Deeks 2005) tenga un poder estadístico menor que las alternativas propuestas recientemente. Esta prueba no se diseñó para el sesgo de publicación en las revisiones sistemáticas de ensayos aleatorios. En su lugar, se concibió como un metanálisis de estudios de exactitud de las pruebas de diagnóstico, donde los odds ratios muy grandes y los estudios con desequilibrios importantes pueden provocar problemas para otras pruebas.

10.4.4 Análisis de sensibilidad

Cuando los revisores encuentran pruebas de efectos de los estudios pequeños, deben considerar los análisis de sensibilidad que examinen cómo cambian los resultados del metanálisis bajo diferentes suposiciones con respecto a los motivos de estos efectos. Se hace énfasis en la naturaleza exploratoria de dichos análisis, debido a la dificultad inherente de ajustar por el sesgo de publicación y la falta de investigaciones sobre el rendimiento de estos métodos aplicados de forma condicional, según los resultados de las pruebas de sesgo de publicación consideradas en la Sección 10.4.3. Esta área está relativamente poco desarrollada, por lo que se han indicado los siguientes enfoques.

10.4.4.1 Comparación de las estimaciones de efectos fijos y aleatorios

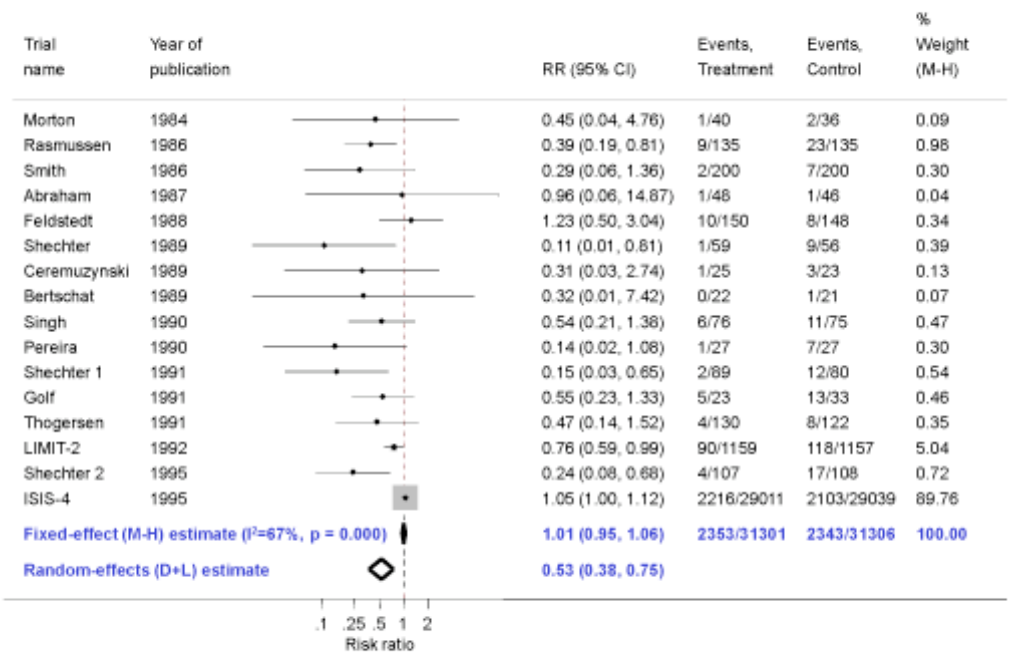
En presencia de heterogeneidad, un metanálisis de efectos aleatorios pondera los estudios de forma relativamente más equitativa los estudios que el análisis de efectos fijos. Por lo tanto, en presencia de efectos de los estudios pequeños como los mostrados en la Figura 10.2.a, en la cual el efecto de la intervención es más beneficioso en los estudios más pequeños, la estimación de efectos aleatorios del efecto de la intervención será más beneficiosa que la de efectos fijos. Poole y Greenland resumieron lo anterior al señalar que “los metanálisis de efectos aleatorios no siempre son conservadores” (Poole 1999). Este aspecto se discute también en el Capítulo 9 (Sección 9.5.4).

Un ejemplo extremo de las diferencias entre los análisis de efectos fijos y aleatorios que pueden surgir en presencia de los efectos de los estudios pequeños se puede observar en la Figura 10.4.c, que muestra las estimaciones para los modelos fijos y aleatorios de los efectos del magnesio intravenoso sobre la mortalidad después del infarto de miocardio. Este es un ejemplo bien conocido en el cual se encontraron efectos beneficiosos de la intervención en un metanálisis de estudios pequeños, que posteriormente se contradijeron cuando un estudio muy grande, el ISIS-4, no encontró pruebas de que el magnesio afectara la mortalidad.

Debido a que existe heterogeneidad significativa entre los ensayos, los estudios tienen una ponderación mucho más equitativa en el análisis de efectos aleatorios que en el análisis de efectos fijos. En el análisis de efectos fijos el ensayo ISIS-4 toma el 90% de la ponderación, por lo que no hay pruebas de un efecto beneficioso de la intervención. En el análisis de efectos aleatorios dominan los estudios pequeños y parece haber pruebas claras de un efecto beneficioso de la intervención. Para interpretar las pruebas acumuladas es necesario evaluar la probable validez de las pruebas combinadas de los estudios más pequeños, en comparación con las del ensayo ISIS-4.

Se recomienda que cuando los revisores tengan inquietud acerca de la influencia de los efectos de los estudios pequeños sobre los resultados de un metanálisis en el cual existan pruebas de heterogeneidad ($I^2 > 0$), comparen las estimaciones del efecto de la intervención con los modelos de efectos fijos y aleatorios. Si las estimaciones son similares, entonces cualquier efecto de los estudios pequeños tiene poco efecto sobre la estimación del efecto de la intervención. Si la estimación del modelo de efectos aleatorios es más beneficiosa, los revisores deben considerar si es razonable concluir que la intervención fue más efectiva en los estudios más pequeños. Si los estudios más grandes tienden a ser los que se realizaron con mayor rigor metodológico en circunstancias más típicas del uso de la intervención en la práctica, los revisores deben considerar reportar los resultados de los metanálisis limitados a los estudios más grandes y más rigurosos. Sería conveniente la evaluación formal de estas estrategias en estudios de simulación. Hay que señalar que no son posibles las comparaciones estadísticas formales de las estimaciones del efecto de la intervención mediante los modelos de efectos fijos y aleatorios, y que aún es posible que los efectos de los estudios pequeños sesguen los resultados de un metanálisis en el cual no haya pruebas de heterogeneidad, incluso si las estimaciones del efecto de la intervención mediante los modelos de efectos fijos y aleatorios son idénticas en esta situación.

Figura 10.4.c: Comparación del efecto del magnesio intravenoso sobre la mortalidad tras un infarto de miocardio mediante estimaciones metanalíticas del modelo de efectos fijos de efectos aleatorios



10.4.4.2 Recortar y rellenar

El método de “recortar y rellenar” tiene como objetivo identificar y corregir la asimetría del gráfico de embudo que surge del sesgo de publicación (Taylor 1998, Duval 2000). Las bases del método son (1) “recortar” (eliminar) los estudios más pequeños que provocan la asimetría del gráfico de embudo, (2) utilizar el gráfico recortado para la estimación del “centro” verdadero del embudo, y (3) reemplazar los estudios omitidos y sus “contrapartes” perdidas alrededor del centro (rellenar). Al tiempo que proporciona una estimación del número de estudios que faltan, se deriva un efecto de la intervención ajustado al realizar un metanálisis con la inclusión de los estudios rellenados.

El método de recortar y rellenar no necesita hacer suposiciones acerca de los mecanismos que dan lugar al sesgo de publicación, proporciona una estimación del número de estudios faltantes y también brinda una estimación del efecto de la intervención “ajustada” por el sesgo de publicación (según los estudios rellenados). Sin embargo, el mismo se basa en la suposición sólida de que debe haber un gráfico de embudo simétrico y no garantiza que el efecto de la intervención ajustado sea igual que el que se observaría en ausencia de sesgo de publicación, ya que no es posible conocer el verdadero mecanismo del sesgo de publicación. De similar importancia, el método de recortar y rellenar no toma en cuenta otros motivos para la asimetría del gráfico de embudo diferentes del sesgo de publicación. Por lo tanto, las estimaciones “corregidas” del efecto de la intervención a partir de este método se deben interpretar con gran precaución. Se sabe que el método tiene un rendimiento deficiente en presencia de heterogeneidad significativa entre los estudios (Terrin 2003, Peters 2007). Adicionalmente, la estimación y las inferencias se basan en un conjunto de datos que contiene estimaciones imputadas del efecto de la intervención. Se pudiera argumentar que dichas estimaciones contribuyen de forma inapropiada con información que reduce la incertidumbre en el efecto resumen de la intervención.

10.4.4.3 N a prueba de fallos

Rosenthal sugirió evaluar la posibilidad de que el sesgo de publicación haya influido en los resultados de un metanálisis mediante el cálculo del "N a prueba de fallos", que es el número de estudios "negativos" adicionales (estudios en los cuales el efecto de la intervención fue cero) que se necesitaría para aumentar el valor de P del metanálisis por encima de 0,05 (Rosenthal 1979). Sin embargo, la estimación del N a prueba de fallos es muy dependiente de la media del efecto de la intervención que se supone para los estudios no publicados (Iyengar 1988), y los métodos disponibles dan lugar a estimaciones muy variables del número de estudios adicionales (Becker 2005). El método también se opone al principio de que en la investigación médica en general, y en las revisiones sistemáticas en particular, es necesario concentrarse en el tamaño de la estimación del efecto de la intervención y los intervalos de confianza asociados, en lugar de si el valor de P alcanza un umbral particular arbitrario, aunque también se han propuesto métodos relacionados para los tamaños del efecto (Orwin 1983). Por lo tanto, éste y otros métodos relacionados no se recomiendan para su uso en las revisiones Cochrane.

10.4.4.4 Otros modelos de selección

Otros autores han propuesto métodos más sofisticados que evitan las suposiciones sólidas acerca de la asociación entre el valor de P del estudio y la probabilidad de publicación (Dear 1992, Hedges 1992). Estos métodos se pueden extender a la estimación de los efectos de la intervención, corregidos por la estimación del sesgo de publicación (Vevea 1995). Sin embargo, requieren de un gran número de estudios para que se incluya un rango suficiente de valores de la P de los estudios. También se ha propuesto un enfoque bayesiano en el cual se simulan el número y los resultados de estudios inadvertidos como una forma de corregir las estimaciones del efecto de la intervención para el sesgo de publicación (Givens 1997). Trabajos recientes han examinado la posibilidad de evaluar la solidez en un rango de funciones de ponderación, por lo que evitan la necesidad de un gran número de estudios (Vevea 2005). La complejidad de los métodos estadísticos y el gran número de estudios necesarios explican probablemente por qué los modelos de selección no se utilizan ampliamente en la práctica.

10.4.4.5 Análisis de sensibilidad según los modelos de selección

Copas desarrolló un modelo en el cual la probabilidad de que un estudio se incluya en un metanálisis depende de su error estándar. Debido a que no es posible realizar la estimación precisa de todos los parámetros del modelo, el autor recomienda análisis de sensibilidad en los cuales el valor de la estimación del efecto de la intervención se computa bajo un rango de suposiciones acerca de la intensidad del sesgo de selección (Copas 1999). En lugar de una estimación única "corregida" del efecto de la intervención por el sesgo de publicación, el lector puede observar cómo la estimación del efecto (y el intervalo de confianza) varía cuando aumenta la supuesta cantidad de sesgo de selección. La aplicación del método a los estudios epidemiológicos medioambientales del hábito de fumar y cáncer de pulmón indica que el sesgo de publicación puede explicar algunas asociaciones observadas en los metanálisis de estos estudios (Copas 2000).

10.4.4.6 Pruebas para el exceso de estudios con resultados significativos

Ioannidis y Trikalinos proponen una prueba sencilla que tiene como objetivo evaluar si hay un exceso de estudios que tienen formalmente resultados estadísticamente significativos (Ioannidis 2007a). La prueba compara el número de estudios que formalmente tienen resultados estadísticamente significativos con el número de resultados esperados bajo diferentes suposiciones acerca de la magnitud del tamaño del efecto. La suposición más sencilla es que el tamaño del efecto es igual al efecto resumen observado en el metanálisis (pero el mismo puede introducir un elemento de circularidad). Se pueden utilizar otros valores para el tamaño del efecto subyacente, así como diferentes umbrales de significación.

Por lo tanto, al igual que el gráfico de embudo de contorno mejorado descrito en la Sección 10.4.1, pero al contrario de las pruebas de regresión, este método toma en cuenta la distribución de la significación de los resultados del estudio. Sin embargo, a diferencia de las pruebas de regresión y del gráfico de embudo de contorno mejorado, la prueba no hace suposiciones acerca de los efectos de los estudios pequeños. Un exceso de resultados significativos puede reflejar la eliminación de todos los estudios o un análisis selectivo/manipulado relacionado, así como las prácticas de reporte que provocarían un exceso similar.
La prueba tiene un poder estadístico limitado, como la mayoría de las otras pruebas, cuando hay muy pocos estudios y cuando hay pocos estudios con resultados significativos. Debido a que la prueba no se ha evaluado rigurosamente mediante simulación, ni comparado con pruebas alternativas y bajo otros escenarios, actualmente no se recomienda como alternativa a las descritas en la Sección 10.4.3

Un aspecto novedoso de la prueba es que se puede aplicar a un gran número de metanálisis en el mismo campo de investigación para examinar el grado de sesgo de publicación y de reporte selectivo en todos los dominios de la investigación clínica. Nuevamente, sería conveniente hacer evaluaciones adicionales de este enfoque.

10.4.5 Resumen

Aunque existen pruebas claras de que los sesgos de publicación y otros tipos de sesgos de reporte dan lugar a estimaciones exageradamente optimistas de los efectos de la intervención, la detección, corrección y solución del sesgo de publicación es problemática. Las búsquedas exhaustivas son importantes, particularmente para identificar estudios bien definidos como los ensayos aleatorios. Sin embargo, las búsquedas exhaustivas no son suficientes para prevenir algunos potenciales sesgos significativos.

El sesgo de publicación se debe considerar una de varias causas posibles de “efectos de los estudios pequeños” (una tendencia a que las estimaciones del efecto de la intervención sean más beneficiosas en los estudios más pequeños). Los gráficos de embudo permiten a los revisores hacer una evaluación visual de si los efectos de los estudios pequeños pudieran estar presentes en un metanálisis. Para los resultados continuos (numéricos) con efectos de la intervención medidos como diferencias de medias son válidos los gráficos de embudo y las pruebas estadísticas para la asimetría de un gráfico de embudo. Sin embargo, para los resultados dicotómicos con efectos de la intervención expresados como odds ratios, el error estándar del logaritmo del odds ratio está matemáticamente relacionado con el tamaño del odds ratio, incluso en ausencia de los efectos de los estudios pequeños. Lo anterior puede dar lugar a que los gráficos de embudo donde se utilizan los odds ratios (o los odds ratios en una escala logarítmica) parezcan asimétricos y puede provocar que los valores de P de la prueba de Egger y cols. sean demasiado pequeños. Para otras medidas de resultado aún no existe una recomendación firme. Se recomiendan tres pruebas estadísticas para los efectos de los estudios pequeños para su uso en las revisiones Cochrane, siempre que haya al menos 10 estudios. Sin embargo, ninguna está implementada en RevMan y habitualmente se necesita apoyo estadístico. Sólo una de las pruebas ha mostrado funcionar bien cuando la varianza de la heterogeneidad entre los estudios excede 0,1. Los resultados de las pruebas de asimetría del graficó de embudo se deben interpretar con precaución. Cuando existen pruebas de efectos de los estudios pequeños, se debe considerar el sesgo de publicación como una de varias explicaciones posibles. En estos casos, los revisores deben intentar determinar la fuente de los efectos de los estudios pequeños y considerar sus implicaciones en los análisis de sensibilidad.

10.5 Información del capítulo

Editores: Jonathan AC Sterne, Matthias Egger y David Moher del Grupo Cochrane de Métodos de Sesgo.
La versión en inglés de este capítulo debe citarse como: Sterne JAC, Egger M, Moher D (editors). Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Intervention.* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.
Autores que contribuyeron: James Carpenter, Matthias Egger, Roger Harbord, Julian Higgins, David Jones, David Moher, Jonathan Sterne, Alex Sutton y Jennifer Tetzlaff.
Agradecimientos: Se agradecen a Doug Altman, Jon Deeks, John Ioannidis, Jaime Peters y Gerta Rücker sus útiles comentarios.
Declaraciones de interés: James Carpenter, Jon Deeks, Matthias Egger, Roger Harbord, David Jones, Jaime Peters, Gerta Rücker, Jonathan Sterne y Alex Sutton son autores de artículos que proponen pruebas para la simetría de los gráficos de embudo.

Recuadro 10.5.a: El Grupo Cochrane de Métodos de Sesgo

El Grupo Cochrane de Métodos de Sesgo (GMS), anteriormente llamado Grupo de Métodos de Sesgo de Reporte, se registró formalmente como Grupo de Métodos en el año 2000. El GMS aborda varias formas diferentes de sesgo como el sesgo de publicación, el sesgo de idioma, el sesgo de informe selectivo de resultado y los sesgos que surgen del diseño y la realización del estudio. Una iniciativa importante del grupo, en colaboración con el Grupo de Métodos Estadísticos, fue el desarrollo de una nueva guía para evaluar el riesgo de sesgo de los estudios incluidos en las revisiones Cochrane.

Las actividades de los miembros del GMS incluyen:

realizar investigaciones empíricas para examinar si, y en qué circunstancias, los diferentes sesgos pueden tener una repercusión significativa en las revisiones sistemáticas, incluida la preparación de revisiones Cochrane de Metodología;

realizar investigaciones metodológicas sobre cómo identificar y abordar los posibles sesgos en las revisiones sistemáticas y los metanálisis;

ayudar a completar y coordinar revisiones sistemáticas de Métodos, relevantes para los objetivos del grupo;

proporcionar asesoramiento a las entidades Cochrane; y

ofrecer entrenamiento a autores de revisiones sistemáticas Cochrane y no Cochrane mediante oportunidades formales e informales.

La lista de correo electrónico de los miembros del GMS se utiliza como un forum de discusión y diseminación de información. Los boletines de novedades Cochrane y las listas de distribución por correo electrónico, como los boletines del Grupo de Métodos de Sesgo, la *Cochrane News* y CCInfo también se utilizan para la diseminación de las actividades del grupo.

Financiamiento: El GMS recibe financiamiento de infraestructura como parte de un compromiso de los Canadian Institutes of Health Research (CIHR) y la Canadian Agency for Drugs and Technologies in Health (CADTH) de financiar las entidades Cochrane con base en Canadá. El mismo fondo apoya las actividades de diseminación, el mantenimiento de la web, viajes, entrenamiento, talleres y un puesto de coordinador a tiempo completo.
Sitio web: www.chalmersresearch.com\bmg

10.6 Referencias

Abassi 2004
Abbasi K. Compulsory registration of clinical trials. *BMJ* 2004; 329: 637-638.

Abbot 1998
Abbot NC, Ernst E. Publication bias: direction of outcome is less important than scientific quality. *Perfusion* 1998; 11: 182-182.

Anonymous 1991
Anonymous. Subjectivity in data analysis. *The Lancet* 1991; 337: 401-402.

Bailey 2002
Bailey BJ. Duplicate publication in the field of otolaryngology-head and neck surgery. *Otolaryngology and Head and Neck Surgery* 2002; 126: 211-216.

Barden 2003
Barden J, Edwards JE, McQuay HJ, Moore RA. Oral valdecoxib and injected parecoxib for acute postoperative pain: a quantitative systematic review. *BMC Anesthesiology* 2003; 3: 1.

Bardy 1998
Bardy AH. Bias in reporting clinical trials. *British Journal of Clinical Pharmacology* 1998; 46: 147-150.

Becker 2005
Becker BJ. Failsafe *N* or file-drawer number. In: Rothstein HR, Sutton AJ, Borenstein M (editors). *Publication Bias in Meta-Analysis*. Chichester (UK): John Wiley & Sons, 2005.

Begg 1988
Begg CB, Berlin JA. Publication bias: a problem in interpreting medical data. *Journal of the Royal Statistical Society Series A* 1988; 151: 419-463.

Begg 1994
Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; 50: 1088-1101.

Bhandari 2004
Bhandari M, Busse JW, Jackowski D, Montori VM, Schünemann H, Sprague S, Mears D, Schemitsch EH, Heels-Ansdell D, Devereaux PJ. Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials. *Canadian Medical Association Journal* 2004; 170: 477-480.

Blumenthal 1997
Blumenthal D, Campbell EG, Anderson MS, Causino N, Louis KS. Withholding research results in academic life science. Evidence from a national survey of faculty. *JAMA* 1997; 277: 1224-1228.

Brooks 1985
Brooks TA. Private acts and public objects: an investigation of citer motivations. *Journal of the American Society for Information Science* 1985; 36: 223-229.

Burdett 2003
Burdett S, Stewart LA, Tierney JF. Publication bias and meta-analyses: a practical example. *International Journal of Technology Assessment in Health Care* 2003; 19: 129-134.

Cantekin 1991
Cantekin EI, McGuire TW, Griffith TL. Antimicrobial therapy for otitits media with effusion ('secretory' otitits media). *JAMA* 1991; 266: 3309-3317.

Carter 2006
Carter AO, Griffin GH, Carter TP. A survey identified publication bias in the secondary literature. *Journal of Clinical Epidemiology* 2006; 59: 241-245.

Chan 2004a
Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004; 291: 2457-2465.

Chan 2004b
Chan AW, Krleža-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *Canadian Medical Association Journal* 2004; 171: 735-740.

Chan 2005
Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005; 330: 753.

CLASP Collaborative Group 1994
CLASP Collaborative Group. CLASP: a randomized trial of low-dose aspirin for the prevention and treatment of pre-eclampsia among 9364 pregnant women. *The Lancet* 1994; 343: 619-629.

Cook 1993
Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A, McIlroy W, Oxman AD. Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA* 1993; 269: 2749-2753.

Copas 1999
Copas J. What works?: selectivity models and meta-analysis. *Journal of the Royal Statistical Society Series A* 1999; 162: 95-109.

Copas 2000
Copas JB, Shi JQ. Reanalysis of epidemiological evidence on lung cancer and passive smoking. *BMJ* 2000; 320: 417-418.

Cowley 1993
Cowley AJ, Skene A, Stainer K, Hampton JR. The effect of lorcainide on arrhythmias and survival in patients with acute myocardial infarction: an example of publication bias. *International Journal of Cardiology* 1993; 40: 161-166.

Davey Smith 1994
Davey Smith G, Egger M. Who benefits from medical interventions? Treating low risk patients can be a high risk strategy. *BMJ* 1994; 308: 72-74.

Dear 1992
Dear KBG, Begg CB. An approach to assessing publication bias prior to performing a meta-analysis. *Statistical Science* 1992; 7: 237-245.

Decullier 2005
Decullier E, Lheritier V, Chapuis F. Fate of biomedical research protocols and publication bias in France: retrospective cohort study. *BMJ* 2005; 331: 19.

Decullier 2007
Decullier E, Chapuis F. Oral presentation bias: a retrospective cohort study. *Journal of Epidemiology and Community Health* 2007; 61: 190-193.

Deeks 2005
Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology* 2005; 58: 882-893.

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: 374-378.

Dickersin 1993
Dickersin K, Min YI. NIH clinical trials and publication bias. *Online Journal of Current Clinical Trials* 1993; Doc No 50.

Dickersin 1994
Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; 309: 1286-1291.

Dickersin 1997
Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educucation and Prevention* 1997; 9: 15-21.

Dickersin 2002
Dickersin K, Olson CM, Rennie D, Cook D, Flanagin A, Zhu Q, Reiling J, Pace B. Association between time interval to publication and statistical significance. *JAMA* 2002; 287: 2829-2831.

Dong 1997
Dong BJ, Hauck WW, Gambertoglio JG, Gee L, White JR, Bubp JL, Greenspan FS. Bioequivalence of generic and brand-name levothyroxine products in the treatment of hypothyroidism [see comments]. *JAMA* 1997; 277: 1205-1213.

Duval 2000
Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000; 56: 455-463.

Easterbrook 1991
Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *The Lancet* 1991; 337: 867-872.

Egger 1997a
Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315: 629-634.

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

Egger 2003
Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessment* 2003; 7: 1.

Epstein 1990
Epstein WM. Confirmational response bias among social work journals. *Science, Technology and Human Values* 1990; 15: 9-37.

Ernst 1994
Ernst E, Resch KL. Reviewer bias: A blinded experimental study. *Journal of Laboratory and Clinical Medicine* 1994; 124: 178-182.

Fergusson 2000
Fergusson D, Laupacis A, Salmi LR, McAlister FA, Huet C. What should be included in meta-analyses? An exploration of methodological issues using the ISPOT meta-analyses. *International Journal of Technology Assessment in Health Care* 2000; 16: 1109-1119.

Galandi 2006
Galandi D, Schwarzer G, Antes G. The demise of the randomised controlled trial: bibliometric study of the German-language health care literature, 1948 to 2004. *BMC Medical Research Methodology* 2006; 6: 30.

Givens 1997
Givens GH, Smith DD, Tweedie RL. Publication bias in meta-analysis:a Bayesian data-augmentation approach to account for issues exemplified in the passive smoking debate. *Statistical Science* 1997; 12: 221-250.

Glasziou 1995
Glasziou PP, Iriwg LM. An evidence based approach to individualising treatment. *BMJ* 1995; 311: 1356-1359.

Godlee 1999
Godlee F, Dickersin K. Bias, subjectivity, chance, and conflict of interest in editorial decisions. In: Godlee F, Jefferson T (editors). *Peer Review in Health Sciences*. London (UK): BMJ Books, 1999.

Gøtzsche 1987
Gøtzsche PC. Reference bias in reports of drug trials. *British Medical Journal (Clinical Research Edition)* 1987; 295: 654-656.

Gøtzsche 1989
Gøtzsche PC. Multiple publication of reports of drug trials. *European Journal of Clinical Pharmacology* 1989; 36: 429-432.

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? *Journal of Clinical Epidemiology* 1995; 48: 159-163.

Harbord 2006
Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine* 2006; 25: 3443-3457.

Hartling 2004
Hartling L, Craig WR, Russell K, Stevens K, Klassen TP. Factors influencing the publication of randomized controlled trials in child health research. *Archives of Pediatrics and Adolescent Medicine* 2004; 158: 983-987.

Hedges 1992
Hedges LV. Modeling publication selection effects in meta-analysis. *Statistical Science* 1992; 7: 246-255.

Hemminki 1980
Hemminki E. Study of information submitted by drug companies to licensing authorities. *British Medical Journal* 1980; 280: 833-836.

Heres 2006
Heres S, Davis J, Maino K, Jetzinger E, Kissling W, Leucht S. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *American Journal of Psychiatry* 2006; 163: 185-194.

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: 374-380.

Hopewell 2004
Hopewell S. *Impact of grey literature on systematic reviews of randomized trials* (PhD thesis). University of Oxford, 2004.

Hopewell 2007a
Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art No: MR000011.

Hopewell 2007b
Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art No: MR000010.

Hopewell 2008
Hopewell S, Louden K, Clarke M, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of SystematicReviews* (to appear).

Huston 1996
Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *The Lancet* 1996; 347: 1024-1026.

Hutchison 1995
Hutchison BG, Oxman AD, Lloyd S. Comprehensiveness and bias in reporting clinical trials. *Canadian Family Physician* 1995; 41: 1356-1360.

Ioannidis 1998
Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998; 279: 281-286.

Ioannidis 2001
Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001; 285: 437-443.

Ioannidis 2007a
Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. *Clinical Trials* 2007; 4: 245-253.

Ioannidis 2007b
Ioannidis JP, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Canadian Medical Association Journal* 2007; 176: 1091-1096.

Irwig 1998
Irwig L, Macaskill P, Berry G, Glasziou P. Bias in meta-analysis detected by a simple, graphical test. Graphical test is itself biased. *BMJ* 1998; 316: 470-471.

Iyengar 1988
Iyengar S, Greenhouse JB. Selection problems and the file drawer problem. *Statistical Science* 1988: 109-135.

Johansen 1999
Johansen HK, Gøtzsche PC. Problems in the design and reporting of trials of antifungal agents encountered during meta-analysis [see comments]. *JAMA* 1999; 282: 1752-1759.

Jüni 2002
Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *International Journal of Epidemiology* 2002; 31: 115-123.

Kjaergard 2002
Kjaergard LL, Gluud C. Citation bias of hepato-biliary randomized clinical trials. *Journal of Clinical Epidemiology* 2002; 55: 407-410.

Lexchin 2003
Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003; 326: 1167-1170.

Liebeskind 2006
Liebeskind DS, Kidwell CS, Sayre JW, Saver JL. Evidence of publication bias in reporting acute stroke clinical trials. *Neurology* 2006; 67: 973-979.

Light 1984
Light RJ, Pillemer DB. *Summing up. The science of reviewing research* (1). Cambridge (MA): Harvard University Press, 1984.

Macaskill 2001
Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine* 2001; 20: 641-654.

Mahoney 1977
Mahoney MJ. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research* 1977; 1: 161-175.

Mandel 1987
Mandel EH, Rockette HE, Bluestone CD, Paradise JL, Nozza RJ. Efficacy of amoxicillin with and without decongestant-antihistamine for otitis media with effusion in children. *New England Journal of Medicine* 1987; 316: 432-437.

McAuley 2000
McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *The Lancet* 2000; 356: 1228-1231.

Melander 2003
Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine - selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003; 326: 1171-1173.

Moher 1996
Moher D, Fortin P, Jadad AR, Jüni P, Klassen T, Le Lorier J, Liberati A, Linde K, Penna A. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *The Lancet* 1996; 347: 363-366.

Moher 2000
Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A. What contributions do languages other than English make on the results of meta-analyses? *Journal of Clinical Epidemiology* 2000; 53: 964-972.

Moher 2003
Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomised trials published in languages other than English in systematic reviews. *Health Technology Assessment* 2003; 7: 1-90.

Moher 2007
Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Medicine* 2007; 4: e78.

Moscati 1994
Moscati R, Jehle D, Ellis D, Fiorello A, Landi M. Positive-outcome bias: comparison of emergency medicine and general medicine literatures. *Academic Emergency Medicine* 1994; 1: 267-271.

Olson 2002
Olson CM, Rennie D, Cook D, Dickersin K, Flanagin A, Hogan JW, Zhu Q, Reiling J, Pace B. Publication bias in editorial decision making. *JAMA* 2002; 287: 2825-2828.

Orwin 1983
Orwin RG. A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics* 1983; 8: 157-159.

Peters 1982
Peters DP, Ceci SJ. Peer review practices of psychology journals: The fate of published articles, submitted again. *The Behavioral and Brain Sciences* 1982; 5: 187-255.

Peters 2006
Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 2006; 295: 676-680.

Peters 2007
Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine* 2007; 26: 4544-4562.

Peters 2008
Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. The contour enhanced funnel plot: an aid to interpreting funnel asymmetry. *Journal of Clinical Epidemiology* 2008; 61: 991-996.

Pham 2005
Pham B, Klassen TP, Lawson ML, Moher D. Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. *Journal of Clinical Epidemiology* 2005; 58: 769-776.

Pittler 2000
Pittler MH, Abbot NC, Harkness EF, Ernst E. Location bias in controlled clinical trials of complementary/alternative therapies. *Journal of Clinical Epidemiology* 2000; 53: 485-489.

Pocock 1987
Pocock S, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *New England Journal of Medicine* 1987; 317: 426-432.

Poole 1999
Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *American Journal of Epidemiology* 1999; 150: 469-475.

Ravnskov 1992
Ravnskov U. Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *BMJ* 1992; 305: 15-19.

Rennie 1991
Rennie D. The Cantekin affair. *JAMA* 1991; 266: 3333-3337.

Rennie 1997
Rennie D. Thyroid Storms. *JAMA* 1997; 277: 1238-1243.

Rosenthal 1979
Rosenthal R. The 'file drawer problem' and tolerance for null results. *Psychological Bulletin* 1979; 86: 638-641.

Rücker 2008
Rücker G, Schwarzer G, Carpenter J. Arcsine test for publication bias in meta-analyses with binary outcomes. *Statistics in Medicine* 2008; 27: 746-763.

Sampson 2003
Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R, St John PD, Viola R, Raina P. Should meta-analysts search Embase in addition to Medline? *Journal of Clinical Epidemiology* 2003; 56: 943-955.

Scherer 2007
Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art No: MR000005.

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: 408-412.

Schwarzer 2007
Schwarzer G, Antes G, Schumacher M. A test for publication bias in meta-analysis with sparse binary data. *Statistics in Medicine* 2007; 26: 721-733.

Simes 1987
Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Statistics in Medicine* 1987; 6: 11-29.

Smith 1999
Smith R. What is publication? A continuum. *BMJ* 1999; 318: 142.

Sterling 1959
Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa. *Journal of the American Statistical Association* 1959; 54: 30-34.

Sterling 1995
Sterling TD, Rosenbaum WL, Weinkam JJ. Publication decisions revisted: The effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statatistician* 1995; 49: 108-112.

Stern 1997
Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997; 315: 640-645.

Sterne 2000
Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology* 2000; 53: 1119-1129.

Sterne 2001
Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology* 2001; 54: 1046-1055.

Stuck 1998
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. *BMJ* 1998; 316: 469-471.

Tang 2000
Tang JL, Liu JL. Misleading funnel plot for detection of bias in meta-analysis. *Journal of Clinical Epidemiology* 2000; 53: 477-484.

Tannock 1996
Tannock IF. False-positive results in clinical trials: multiple significance tests and the problem of unreported comparisons. *Journal of the National Cancer Institute* 1996; 88: 206-207.

Taylor 1998
Taylor SJ, Tweedie RL. Practical estimates of the effect of publication bias in meta-analysis. *Australian Epidemiologist* 1998; 5: 14-17.

Teo 1993
Teo KK, Yusuf S, Furberg CD. Effects of prophylactic antiarrhythmic drug therapy in acute myocardial infarction. An overview of results from randomized controlled trials [see comments]. *JAMA* 1993; 270: 1589-1595.

Terrin 2003
Terrin N, Schmid CH, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine* 2003; 22: 2113-2126.

Terrin 2005
Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *Journal of Clinical Epidemiology* 2005; 58: 894-901.

Tetzlaff 2006
Tetzlaff J, Moher D, Pham B, Altman D. Survey of views on including grey literature in systematic reviews. *14th Cochrane Colloquium*, Dublin (Ireland), 2006.

Tramèr 1997
Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997; 315: 635-640.

Vevea 1995
Vevea JL, Hedges LV. A general linear model for estimating effect size in the presence of publication bias. *Psychometrika* 1995; 60: 419-435.

Vevea 2005
Vevea JL, Woods CM. Publication bias in research synthesis: sensitivity analysis using a priori weight functions. *Psychological Methods* 2005; 10: 428-443.

Vickers 1998
Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Controlled Clinical Trials* 1998; 19: 159-166.

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

Weber 1998
Weber EJ, Callaham ML, Wears RL, Barton C, Young G. Unpublished research from a medical specialty meeting: why investigators fail to publish. *JAMA* 1998; 280: 257-259.

Zarin 2005
Zarin DA, Tse T, Ide NC. Trial Registration at ClinicalTrials.gov between May and October 2005. *New England Journal of Medicine* 2005; 353: 2779-2787.

ÍNDICE
CAPÍTULO 11: PRESENTACIÓN DE LOS RESULTADOS Y LAS TABLAS "RESUMEN DE LOS RESULTADOS"

PUNTOS CLAVE

11.1 INTRODUCCIÓN

11.2 RESULTADOS DE LA BÚSQUEDA Y SELECCIÓN DE ESTUDIOS

11.2.1 Diagrama de flujo de los estudios
Figura 11.2.a: Ejemplo de un diagrama de flujo de un estudio según el modelo PRISMA
11.2.2 Tablas "Características de los estudios incluidos"

11.3 DATOS Y ANÁLISIS

11.3.1 La sección "Datos y análisis" de una revisión
11.3.2 Diagramas de bosque
11.3.2.1 Diagramas de bosque en RevMan
Recuadro 11.3.a: Detalles proporcionados en un diagrama de bosque Cochrane
Figura 11.3.a: Ejemplo de un diagrama de bosque de RevMan
11.3.3. Tablas de otros datos

11.4 FIGURAS

11.4.1 Tipos de figuras
11.4.2 Selección de los análisis de RevMan como figuras
11.4.3 Figuras adicionales

11.5 TABLAS "RESUMEN DE LOS RESULTADOS"

11.5.1 Introducción a las tablas "Resumen de los resultados"
11.5.2 Selección de los resultados para las tablas "Resumen de los resultados"
11.5.3 Plantilla general para las tablas "Resumen de los resultados"
Figura 11.5.a: Ejemplo de una tabla "Resumen de los resultados"
11.5.4 Producción de las tablas "Resumen de los resultados"
11.5.5 Consideraciones estadísticas en las tablas "Resumen de los resultados"
11.5.6 Contenidos detallados de una tabla "Resumen de los resultados"
11.5.6.1 Título y encabezado de la tabla
11.5.6.2 Resultados
11.5.6.3 Riesgos comparativos ilustrativos 1: Riesgo asumido (con la intervención control)
11.5.6.4 Riesgos comparativos ilustrativos 2: Riesgo esperado (con la intervención experimental)
11.5.6.5 Efecto relativo (IC del 95%)
11.5.6.6 Número de participantes (estudios)
11.5.6.7 Calidad de las pruebas (GRADE)
11.5.6.8 Comentarios

11.6 TABLAS ADICIONALES

11.7 PRESENTACIÓN DE LOS RESULTADOS EN EL TEXTO

11.7.1 Resultados de los metanálisis
11.7.2 Resultados sin metanálisis

11.8 REDACCIÓN DE UN RESUMEN

Recuadro 11.8.a: Ejemplo hipotético de un resumen

11.9 REDACCIÓN DE UN RESUMEN EN LENGUAJE SENCILLO

11.9.1 Acerca de los resúmenes en lenguaje sencillo
11.9.2 Título en lenguaje sencillo
11.9.3 Texto resumen

11.10 INFORMACIÓN DEL CAPÍTULO

11.11 REFERENCIAS

CAPÍTULO 11: PRESENTACIÓN DE LOS RESULTADOS Y LAS TABLAS "RESUMEN DE LOS RESULTADOS"

Autores: Holger J Schünemann, Andrew D Oxman, Julian PT Higgins, Gunn E Vist, Paul Glasziou y Gordon H Guyatt del Grupo Cochrane de Métodos de Aplicabilidad y Recomendaciones y del Grupo Cochrane de Métodos Estadísticos.

Puntos clave

- Las tablas y figuras ayudan a presentar los estudios incluidos y sus resultados en un formato sistemático y claro.
- Los diagramas de bosque ('forest plots') son la manera estándar de ilustrar los resultados de los estudios individuales y los metanálisis. Es posible generarlos mediante el programa informático Review Manager y hacer una selección para su inclusión en el cuerpo de una revisión sistemática.
- La tabla "Resumen de los resultados" proporciona información clave con respecto a la calidad de las pruebas, la magnitud del efecto de las intervenciones examinadas y el resumen de los datos disponibles sobre todos los resultados importantes para una comparación determinada.
- El Resumen de una revisión Cochrane debe dirigirse principalmente a los encargados de tomar decisiones sanitarias (incluidos los médicos, los consumidores informados y los responsables de las políticas de salud); y el "Resumen en lenguaje sencillo" transmite los resultados en un estilo directo que pueda comprender el usuario o consumidor de la atención sanitaria.

11.1 Introducción

La sección Resultados de una revisión debe resumir los resultados en un orden claro y lógico, y debe abordar de manera explícita los objetivos de la revisión. Los revisores pueden utilizar varias tablas y figuras para presentar la información en un formato más conveniente:

- Tablas "Características de los estudios incluidos" (incluidas las tablas "Riesgo de sesgo").
- "Datos y análisis" (el conjunto completo de tablas de datos y diagramas de bosque).
- Figuras (una selección de diagramas de bosque, gráficos en embudo, gráficos de "Riesgo de sesgo" y otras figuras).
- Tablas "Resumen de los resultados".
- Tablas adicionales.

Las tablas "Características de los estudios incluidos" presentan la información sobre los estudios individuales; las tablas "Datos y análisis" y los diagramas de bosque presentan los datos de los resultados de los estudios individuales y pueden incluir adicionalmente los metanálisis; las tablas "Resumen de los resultados" presentan la información acumulativa, los datos y la calidad de las pruebas para los resultados más importantes. Los resultados de una revisión también se deben condensar en un resumen y en un resumen en lenguaje sencillo.

Las tablas "Resumen de los resultados" son clave entre estos instrumentos de presentación de la información, y una parte fundamental de este capítulo está dedicado a ellas. Se discute la especificación de los resultados importantes que pudieran ser relevantes para las personas que consideran la/s intervención/es en estudio, un paso que a menudo se subvalora en las revisiones Cochrane. Se presentan ejemplos de tablas "Resumen de los resultados" y se describe el contenido de dichas tablas. El Capítulo 12 discute los aspectos de la interpretación de los resultados.

11.2 Resultados de la búsqueda y selección de estudios

11.2.1 Diagrama de flujo de los estudios

Los diagramas de flujo de los estudios se utilizan para ilustrar los resultados de la búsqueda y el proceso de evaluación y selección de los estudios para su inclusión en la revisión. La figura 11.2.a es un ejemplo de un diagrama de flujo, siguiendo el modelo descrito en la declaración PRISMA (Liberati 2009). Se puede crear un diagrama de flujo según el modelo de PRISMA con RevMan, aunque RevMan también incluye la posibilidad de crear un diagrama de flujo con una estructura flexible.

Es importante que el texto del diagrama de flujo establezca una clara distinción entre los estudios y los registros. Los registros son fuentes de información sobre un estudio, tales como artículos de revistas, capítulos de libros, páginas web y otros documentos. Los estudios son los estudios de investigación propiamente, habitualmente ensayos clínicos aleatorizados en el contexto de una revisión Cochrane. Por lo general, un diagrama de flujo comienza describiendo el número de registros recuperados (la mayoría de los cuales habitualmente a partir de bases de datos bibliográficas). Después de eliminar los duplicados, los registros son relacionados con estudios distintos (véase el capítulo 7, sección 7.2), y el diagrama de flujo debe reflejarlo cambiando su énfasis a los estudios.

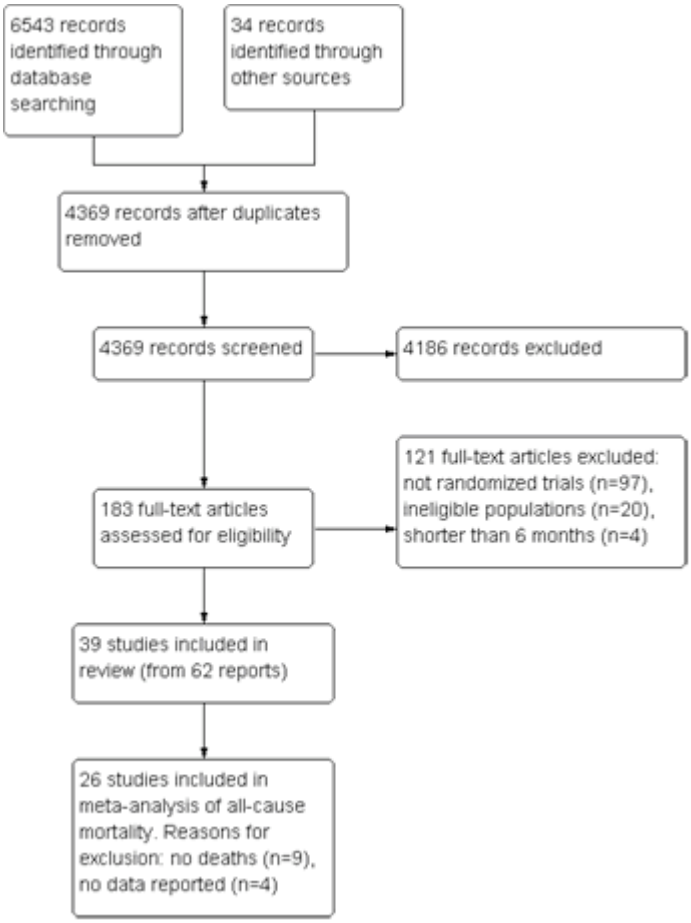
El diagrama de flujo deberá presentar:

- número de registros únicos identificados por las búsquedas;
- número de registros excluidos después de la evaluación preliminar (por ejemplo, a partir de los títulos y resúmenes);
- número de registros recuperados en texto completo;

- número de registros o estudios excluidos después de la evaluación del texto completo, con una breve explicación de los motivos;
- número de estudios que cumplieron con los criterios de elegibilidad para la revisión (y que, por tanto, contribuyen a la síntesis cualitativa), y
- número de estudios que contribuyen al resultado principal.

En el artículo que explica la justificación y el proceso de elaboración de la declaración PRISMA (Liberati 2009) y en una revisión de diagramas de flujo publicados (Hind 2007) se proporciona más detalles sobre la preparación de un diagrama de flujo de un estudio.

Figura 11.2.a: Ejemplo de un diagrama de flujo de un estudio según el modelo PRISMA



11.2.2 Tablas "Características de los estudios incluidos"

Los revisores deben decidir qué características de los estudios es probable que sean relevantes para los usuarios de la revisión. Es probable que la mayoría de estas características hayan sido mencionadas en el protocolo. Los revisores deben, como mínimo, incluir las siguientes en la tabla "Características de los estudios incluidos":

Métodos: diseño del estudio (señalar si el estudio fue aleatorio o no) e indicar claramente, cuando sea relevante, si el estudio difiere de un diseño de grupos paralelos estándar (p.ej. un diseño cruzado o aleatorio grupal o por 'clusters'); la duración del estudio (si no se incluye en el apartado de la Intervención). Nota: el ítem "Métodos" no debe incluir las medidas del riesgo de sesgo, las cuales deben aparecer en la tabla "Riesgo de sesgo" (ver Capítulo 8, Sección 8.5).

Participantes: ámbito; detalles relevantes del estado de salud de los participantes; edad; sexo; país. Se debe proporcionar información suficiente que permita a los usuarios de la revisión determinar la aplicabilidad del estudio a su población, y permitir la exploración de diferencias de los participantes entre los diversos estudios.
Intervención: una lista clara de los grupos de intervención incluidos en cada estudio. De ser factible, se debe proporcionar información suficiente para poder replicar cada intervención en la práctica; para las intervenciones farmacológicas, incluir los detalles del nombre del fármaco, la dosis, frecuencia, forma de administración (si no es evidente), duración (si no se incluyó en Métodos); para las intervenciones no farmacológicas, incluir consideraciones y componentes relevantes relacionados con la intervención.
Resultados: una lista clara de (i) las variables de resultado y puntos temporales del estudio que se consideran en la revisión, o (ii) las variables de resultado y puntos temporales medidos (o informados) en el estudio. Los resultados propiamente del estudio no se deben incluir aquí (ni en otro lugar en esta tabla).
Notas: comentarios adicionales de los revisores relacionados con aspectos del estudio no cubiertos en las categorías anteriores. Es de señalar que las evaluaciones del riesgo de sesgo se deben hacer en la tabla "Riesgo de sesgo".

Es posible agregar hasta tres campos extra en la tabla "Características de los estudios incluidos". De ser apropiado, se recomienda a los revisores que utilicen un campo extra para proporcionar información acerca del financiamiento de cada estudio.

11.3 Datos y análisis

11.3.1 La sección "Datos y análisis" de una revisión

La sección "Datos y análisis" de una revisión Cochrane es una fuente detallada de los resultados. Incluye los datos de los resultados (numéricos o de texto), los diagramas de bosque y los resultados del metanálisis. Lo fundamental del recurso "Datos y análisis" es una tabla de comparaciones, resultados y (opcionalmente) subgrupos para los cuales los datos están disponibles. Los análisis enumerados en esta tabla incluyen una tabla de resultados (tablas "otros datos") o, con mayor frecuencia, una tabla de datos acompañada por un diagrama de bosque. Las tablas "Datos y análisis" se incluyen en la publicación completa de una revisión Cochrane. Sin embargo, algunos formatos de una revisión publicada pueden omitir los diagramas de bosque y las tablas "otros datos" (junto con los apéndices), y por lo tanto, se deben considerar generalmente como material complementario y los resultados clave se deben incluir en el texto de la revisión en la sección "Resultados". La revisión publicada incluirá siempre una tabla resumen de todos los análisis (incluido el número de estudios y los resultados del metanálisis para cada subgrupo en cada resultado y para cada comparación). La revisión debe incluir los diagramas de bosque más importantes del recurso "Datos y análisis" como figuras y los mismos se deben citar en la sección "Resultados" (ver Sección 11.4.2).

11.3.2 Diagramas de bosque

Un diagrama de bosque muestra las estimaciones del efecto y los intervalos de confianza para los estudios individuales y los metanálisis (Lewis 2001). Cada estudio se representa por un bloque en la estimación puntual del efecto de la intervención con una línea horizontal que se extiende a cada lado del bloque. El área del bloque indica la ponderación asignada a ese estudio en el metanálisis, mientras que la línea horizontal representa el intervalo de confianza (habitualmente con un nivel de confianza del 95%). El área del bloque y el intervalo de confianza transmiten información similar, pero ambos hacen contribuciones diferentes al gráfico. El intervalo de confianza representa el rango de los efectos de la intervención compatibles con el resultado del estudio e indica si cada uno de ellos fue estadísticamente significativo desde el punto de vista individual. El tamaño del bloque llama la atención hacia los estudios con una mayor ponderación (habitualmente los que tienen intervalos de confianza más estrechos), que dominan el cálculo del resultado combinado.

11.3.2.1 Diagramas de bosque en RevMan

RevMan proporciona un marco flexible para producir diagramas de bosque en la sección "Datos y análisis" de una revisión Cochrane. Los componentes de un diagrama de bosque Cochrane se describen en el Recuadro 11.3.a, y se proporciona un ejemplo de RevMan en la Figura 11.3.a, utilizando los resultados de una revisión sobre las medias compresivas para prevenir la trombosis venosa profunda en pasajeros de aerolíneas (Clarke 2006). Hay un tutorial sobre el uso de RevMan dentro de dicho programa informático (disponible en www.cc-ims.net).

RevMan ofrece múltiples opciones para cambiar los métodos de análisis (p.ej. entre los metanálisis de efectos fijos y aleatorios), o para utilizar diferentes medidas del efecto, ver Capítulo 9 (Sección 9.4) y gráficos (p.ej. escalas de los ejes y orden de los estudios). Para cada conjunto de datos ingresado en RevMan se incorpora automáticamente un diagrama de bosque en la versión completa publicada de la revisión Cochrane. Los análisis se muestran por defecto, a menos que las opciones se invaliden. Las medidas del efecto utilizadas por defecto son los odds ratios de Mantel-Haenszel para los datos dicotómicos, los metanálisis de efectos fijos de las diferencias de medias para los datos continuos, los odds ratios de Peto para los resultados "O – E y Varianza" y los metanálisis de efectos fijos para los resultados de la varianza inversa genérica (ver Capítulo 9, Sección 9.4). El revisor debe invalidar cualquier configuración por defecto que no corresponda con los resultados informados en el texto cuando configure o edite los resultados en RevMan. Lo anterior asegura que los resultados mostrados sean consistentes con lo que se describe en el texto. Además, la escala de los ejes se debe seleccionar de manera que las estimaciones puntuales (y la mayoría, si no todos, de los intervalos de confianza) sean visibles en el gráfico.

Una convención anterior adoptada en la *Base de Datos Cochrane de Revisiones Sistemáticas* (en inglés, *CDSR*) ha sido que los resultados dicotómicos se centren en los resultados desfavorables, de manera que los cocientes de riesgos y los odds ratios menores de uno (y las diferencias de riesgos menores de cero) indiquen que una intervención experimental es superior a una intervención control. Lo anterior daría lugar a estimaciones del efecto hacia la izquierda de la línea vertical en un diagrama de bosque, lo que implica beneficio de la intervención experimental. Ya no se recomienda esta convención porque no es apropiada para todos los casos. Un enfoque muy superior es establecer con claridad qué lado de la línea indica beneficio de qué intervención al rotular las direcciones en los ejes de los diagramas de bosque. RevMan permite a los revisores especificar los rótulos utilizados para los grupos "experimental" y "control" en cada resultado. Estos rótulos se utilizan luego en la *CDSR*. Por lo tanto, es fundamental conocer en qué dirección se construyen y se deben interpretar las figuras. Lo anterior es particularmente importante para los datos de escalas de medición donde no siempre es evidente para el lector qué dirección en una escala indica un empeoramiento de la salud.

No se deben generar diagramas de bosque que no contengan estudios, y no se recomiendan cuando sólo se encuentra un único estudio para un resultado particular. Para mostrar los resultados que sólo se han investigado en estudios únicos, los revisores pueden utilizar un diagrama de bosque que use un subgrupo para cada resultado (y asegurar que la opción para combinar los datos está deshabilitada). Por otra parte, los resultados de los estudios únicos se pueden presentar de manera más conveniente en una tabla Adicional (ver Sección 11.6).

Recuadro 11.3.a: Detalles proporcionados en un diagrama de bosque Cochrane

Los diagramas de bosque para resultados dicotómicos y resultados expresados como "O – E y Varianza" muestran, por defecto:

1. los datos brutos (correspondientes a la tabla de 2x2) para cada estudio;
2. las estimaciones puntuales y los intervalos de confianza para la medida de efecto seleccionada, en forma de bloques y líneas y también como texto;
3. un metanálisis para cada subgrupo utilizando la medida del efecto y el método de análisis seleccionados (efectos fijos o aleatorios), como un diamante y como texto;
4. el número total de participantes y el número total con eventos en los grupos de intervención experimental e intervención control;
5. estadísticos de la heterogeneidad (varianza entre estudios [tau cuadrado, o Tau^2, o τ^2] para los metanálisis de efectos aleatorios, la prueba de ji cuadrado, el estadístico I^2 y una prueba para las diferencias entre subgrupos si se ha hecho análisis de subgrupos y es apropiado);
6. una prueba para el efecto global (efecto promedio global para los metanálisis de efectos aleatorios); y
7. ponderaciones porcentuales (pesos) asignadas a cada estudio.

Es de señalar que 3 – 7 no se muestran a menos que los datos se hayan combinado. Además, la prueba para las diferencias entre los subgrupos no se muestra para los análisis de Mantel-Haenszel. Para los resultados expresados como "O – E y Varianza" también es posible habilitar que se muestren los estadísticos O – E y V.

Los diagramas de bosque para los resultados continuos muestran, por defecto:

1. los datos brutos (medias, desviaciones estándar y tamaños de muestra) para cada brazo en cada estudio;
2. las estimaciones puntuales y los intervalos de confianza para la medida del efecto seleccionada, en forma de bloques y líneas y también como texto;
3. un metanálisis para cada subgrupo utilizando la medida del efecto y el método de análisis seleccionados (efectos fijos o aleatorios), como un diamante y como texto;
4. el número total de participantes en los grupos experimental y control;
5. estadísticos de la heterogeneidad (varianza entre estudios [tau cuadrado] para los metanálisis de efectos aleatorios, la prueba de ji cuadrado, el estadístico I^2 y una prueba para las diferencias entre subgrupos si se ha realizado análisis de subgrupos);
6. una prueba para el efecto global (efecto promedio global para los metanálisis de efectos aleatorios); y
7. ponderaciones porcentuales (pesos) asignadas a cada estudio.

Es de señalar que 3 – 7 no se muestran a menos que los datos se hayan combinado.

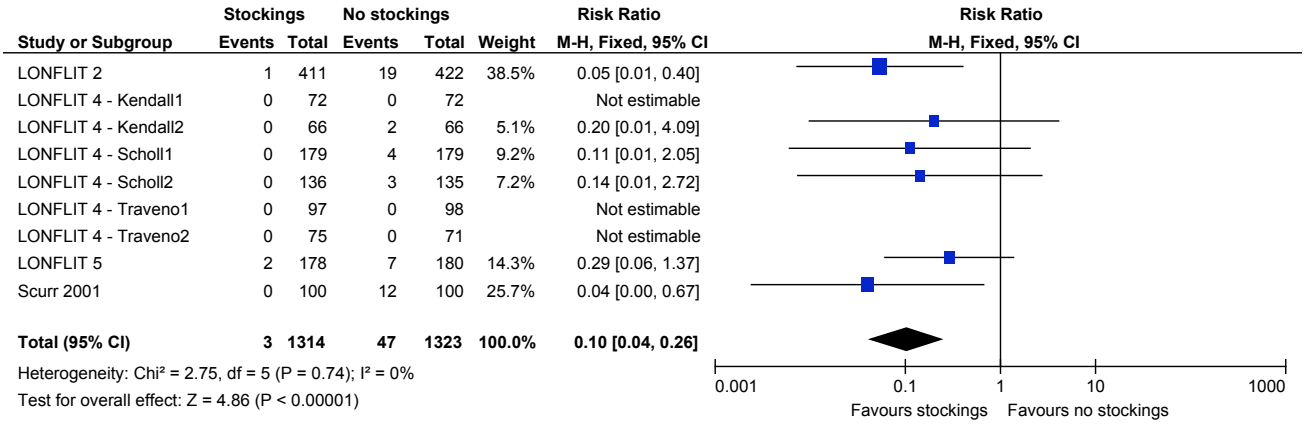
Los diagramas de bosque para el método de la varianza inversa genérica muestran, por defecto:

1. los datos resumen para cada estudio, tal como los ingresó el revisor (las medidas de cocientes [ratios] estarán en la escala logarítmica natural ["ln"]);
2. las estimaciones puntuales y los intervalos de confianza, en forma de bloques y líneas y también como texto (las medidas de cocientes [ratios] estarán en la escala natural en lugar de la escala logarítmica);
3. un metanálisis para cada subgrupo donde se utiliza el método seleccionado (efectos fijos o aleatorios), como un diamante y como texto;
4. estadísticos de la heterogeneidad (varianza entre estudios [tau cuadrado] para los metanálisis de efectos aleatorios, la prueba de ji cuadrado, el estadístico I^2 y una prueba para las diferencias entre subgrupos si se ha realizado un análisis de subgrupos);
5. una prueba para el efecto global (efecto promedio global para los metanálisis de efectos aleatorios); y
6. ponderaciones porcentuales (pesos) asignadas a cada estudio.

Es de señalar que 3 – 7 no se muestran a menos que los datos se hayan combinado. Es posible ingresar adicionalmente los tamaños de muestra de los grupos experimental y control. Los mismos se deben ingresar de manera apropiada para el diseño del estudio. Los tamaños de muestra no se incluyen en el análisis, pero si se ingresan se muestran como:

7. el número de participantes en los grupos experimental y control para cada estudio; y
8. el número total de participantes en los grupos experimental y control.

Figura 11.3.a: Ejemplo de un diagrama de bosque de RevMan



11.3.3. Tablas de otros datos

La sección "Datos y análisis" permite un tipo de resultado "Otros datos". Los resultados de los ensayos individuales se pueden ingresar aquí como texto simple. Esta opción se ajusta bien para el ingreso de estadísticos resumen no estándar como los valores de la mediana, o para los datos básicos que fundamentan las estimaciones y los errores estándar que se han ingresado bajo el tipo de resultado varianza inversa genérica (por ejemplo, las medias y las desviaciones estándar de los ensayos cruzados).

11.4 Figuras

11.4.1 Tipos de figuras

Es posible incluir tres tipos de figuras en el contenido principal de una revisión Cochrane.

1. Diagramas de bosque (ver Sección 11.3.2), de la colección completa de "Datos y análisis" en RevMan.

Gráficos en embudo (ver Capítulo 10, Sección 10.4.1), de la colección completa de "Datos y análisis" en RevMan.

Figuras adicionales.

Debido a que no es posible incluir la sección "Datos y análisis" en algunos formatos publicados de una revisión Cochrane, los revisores deben incorporar los diagramas de bosque más importantes como figuras dentro del cuerpo principal de la revisión, y referirse a ellos en momentos relevantes del texto. Sin embargo, es de señalar que los resultados del metanálisis y del análisis de subgrupos de todos los diagramas de bosque de "Datos y análisis" se incluirán como una tabla en todos los formatos publicados de una revisión Cochrane.

Como regla general, las figuras ofrecen una forma sistemática y clara de presentar los resultados de los estudios individuales y los metanálisis. Sin embargo, a menudo es difícil seguir las revisiones que contienen grandes números de figuras, especialmente cuando cada figura contiene muy poca información. Muchas revistas científicas limitan el número de figuras en un artículo a aproximadamente media docena, y la mayoría de los Grupos Cochrane de Revisión aplican las mismas consideraciones.

Los resultados importantes de todas las figuras se deben analizar de forma general en la sección Resultados del texto de la revisión. Cada vez que en el texto de la revisión se informen resultados numéricos tomados de una figura los revisores deben dejar claro su significado y su derivación, y proporcionar una referencia de la figura relevante.

11.4.2 Selección de los análisis de RevMan como figuras

Los diagramas de bosque y los gráficos en embudo de "Datos y análisis" se deben seleccionar como figuras para que aparezcan como una parte integrada de la revisión Cochrane publicada. Los diagramas de bosque que detallan todos los estudios y los datos de los estudios para los resultados primarios se incluyen habitualmente como figuras. Si existen estudios suficientes, un gráfico en embudo para uno o más de los resultados primarios puede ser un complemento útil para estos diagramas de bosque (ver Capítulo 10, Sección 10.4.1).

11.4.3 Figuras adicionales

Aunque RevMan puede producir diagramas de bosque y gráficos en embudo, puede ser apropiado incluir otros tipos de figuras en una revisión. Por ejemplo:

1. diagramas de bosque "generales", donde cada línea representa un metanálisis en lugar de un estudio (por ejemplo, para ilustrar múltiples análisis de subgrupos o de sensibilidad);
gráficos que ilustren análisis de metarregresión; y
gráficos L'Abbé.

Estos gráficos se pueden producir en programas informáticos diferentes de RevMan e incluirse como una figura "adicional". De la misma manera es posible incluir fotografías y diagramas para su uso en otras partes de una revisión Cochrane.

Pocas veces se necesitan figuras adicionales, y no se deberían utilizar para mostrar diagramas de bosque que se pudieran mostrar mediante RevMan. De ser posible, las figuras de deben producir mediante paquetes informáticos estadísticos que producen gráficos apropiados para la publicación como Stata, SAS, SPSS, S-Plus o programas informáticos especializados en metanálisis. Es posible que los programas de hojas de cálculo con un objetivo general no proporcionen una flexibilidad adecuada ni produzcan resultados de buena calidad.

Existe otro documento, *Considerations and recommendations for figures in Cochrane reviews: Graphs of statistical data* (*Consideraciones y recomendaciones para las figuras en las revisiones Cochrane: Gráficos de datos estadísticos*) que proporciona una guía amplia sobre el contenido de las figuras adicionales que ilustran los datos numéricos (disponible en www.cochrane.org/resources/handbook). El documento incluye descripciones y recomendaciones para los gráficos enumerados anteriormente y para muchos otros. Los revisores deben consultar este documento antes de enviar una revisión que contenga figuras adicionales. Un editor o un asesor estadístico deben evaluar todas las figuras adicionales antes de enviar una revisión Cochrane a la *CDSR*. Los revisores deben conocer que las figuras adicionales a menudo pueden ser grandes y necesitar espacio de almacenamiento valioso en la Biblioteca Cochrane. Existen guías disponibles sobre aspectos técnicos de las figuras adicionales en la documentación de RevMan en http://www.cc-ims.net.

La posibilidad de incorporar figuras adicionales en RevMan técnicamente permite al revisor adjuntar tablas como archivos de gráficos. No se recomienda que los revisores hagan esto debido al gran volumen de espacio de almacenamiento que necesitan los archivos de gráficos. En su lugar, se les recomienda a los revisores que utilicen la función Tablas adicionales, proporcionada para este fin.

11.5 Tablas "Resumen de los resultados"

11.5.1 Introducción a las tablas "Resumen de los resultados"

Las tablas "Resumen de los resultados" presentan los resultados principales de una revisión en un formato tabular claro y sencillo. En particular, proporcionan información clave con respecto a la calidad de las pruebas, la magnitud del efecto de las intervenciones examinadas y el resumen de los datos disponibles sobre los resultados principales. Es de esperar que la mayoría de las revisiones tengan una tabla única de "Resumen de los resultados". Otras revisiones pueden incluir más de una, por ejemplo si la revisión aborda más de una comparación importante, o poblaciones significativamente diferentes. En la *CDSR,* la tabla principal "Resumen de los resultados" de una revisión aparecerá al principio, antes de la sección Antecedentes. Otras tablas "Resumen de los resultados" aparecerán entre las secciones Resultados y Discusión.

La planificación de la tabla "Resumen de los resultados" se realiza de forma temprana en la revisión sistemática, junto con la selección de los resultados que se incluirán en (i) la revisión y (ii) la tabla "Resumen de los resultados". Debido a que este es un paso fundamental que habitualmente no se aborda de manera formal en las revisiones Cochrane tradicionales, aquí se examinarán los aspectos de la selección de los resultados.

11.5.2 Selección de los resultados para las tablas "Resumen de los resultados"

Las revisiones Cochrane comienzan con el desarrollo de una pregunta de revisión y la enumeración de todos los resultados principales que son importantes para los pacientes y otros responsables de tomar decisiones (ver Capítulo 5, Sección 5.4) para asegurar la producción de información óptimamente útil. La consulta y la retroalimentación de un protocolo de revisión pueden mejorar este proceso.

Es probable que los resultados importantes incluyan eventos muy conocidos como mortalidad y morbilidad importante (como los accidentes cerebrovasculares y el infarto de miocardio). Sin embargo, también pueden representar efectos secundarios menores frecuentes o graves poco frecuentes, síntomas y calidad de vida, cargas asociadas con el tratamiento y aspectos relacionados con el uso de los recursos (costos). Las cargas incluyen las demandas de cumplir con una intervención que puede disgustar a los pacientes o los cuidadores (p.ej. la familia), como tenerse que realizar pruebas más frecuentes, o las limitaciones en el estilo de vida que requieren algunas intervenciones.

Con frecuencia, cuando se formulan preguntas que incluyen todos los resultados importantes para los pacientes para la toma de decisiones, los revisores enfrentarán el hecho de que los informes de los ensayos aleatorios no han incluido todos estos resultados. Lo anterior es particularmente cierto para los resultados adversos. Por ejemplo, los ensayos aleatorios pudieran contribuir con datos sobre los efectos previstos y sobre los efectos secundarios frecuentes relativamente menores, pero no abordan el riesgo relativo de resultados adversos poco frecuentes como los intentos de suicidio. El Capítulo 14 discute las estrategias para abordar de manera adecuada los efectos adversos. Para obtener los datos de todos los resultados importantes puede ser necesario examinar los resultados de estudios observacionales.

Si una revisión incluye solamente ensayos aleatorios, pudiera no ser posible abordar todos los resultados importantes dentro de las limitaciones de la revisión. Los revisores deben reconocer estas limitaciones, y hacer que sean claras para los lectores.

Los revisores que asumen el desafío de compilar y resumir las mejores pruebas para todos los resultados relevantes pueden enfrentar varios retos. Los mismos incluyen el hecho de que el análisis de los daños se puede realizar en estudios donde los participantes difieren de los incluidos en los estudios utilizados en el análisis de los beneficios. Por lo tanto, los revisores necesitarán considerar en qué medida, si es que existe, los participantes de los estudios observacionales difieren de los que participan en los ensayos aleatorios. Lo anterior

puede influir en la calidad de las pruebas debido a las inquietudes acerca de la direccionalidad (ver Capítulo 12, Sección 12.2). Cuando no se incluye información sobre estos resultados importantes en la revisión, los revisores deben decirlo. En el Capítulo 13 se discuten adicionalmente estos aspectos.

11.5.3 Plantilla general para las tablas "Resumen de los resultados"

Aunque puede haber buenos motivos para modificar el formato de una tabla "Resumen de los resultados" en algunas revisiones, se ha desarrollado un formato estándar para las mismas con el objetivo de asegurar la consistencia y la facilidad de su uso a través de las revisiones, la inclusión de la información más importante necesaria para los responsables de tomar decisiones y la presentación óptima de esta información. Las tablas Cochrane estándar "Resumen de los resultados" incluyen, por lo tanto, los seis elementos siguientes y utilizan un formato fijo (ver Figura 11.5.a).

1. Una lista de todos los resultados importantes, deseables y no deseables.
2. Una medida de la carga típica de estos resultados (p.ej. riesgo ilustrativo, o media ilustrativa, en el grupo de la intervención control).
3. Magnitud absoluta y relativa del efecto (si ambas son apropiadas).
4. Número de participantes y estudios que abordan estos resultados.
5. Una clasificación de la calidad general de las pruebas para cada resultado (que puede variar por resultado).
6. Espacio para comentarios.

Como una medida de la magnitud del efecto, para los resultados dicotómicos la tabla proporcionará habitualmente una medida relativa (p.ej. cociente de riesgos u odds ratio) y las medidas de riesgo absolutas. Para otros tipos de datos se puede proporcionar una medida absoluta sola (como la diferencia de medias para los datos continuos) o una medida relativa sola (p.ej. cociente de riesgos instantáneo para los datos del tipo 'tiempo hasta el evento'). Sin embargo, de ser posible se deben proporcionar medidas relativas y absolutas del efecto. Las revisiones con más de una comparación principal requieren tablas "Resumen de los resultados" por separado para cada comparación. La Figura 11.5.a proporciona un ejemplo de una tabla "Resumen de los resultados".

En la Sección 11.5.6 aparece una descripción detallada de los contenidos de una tabla "Resumen de los resultados".

Figura 11.5.a: Ejemplo de una tabla "Resumen de los resultados"
Resumen de los resultados:

Medias compresivas comparadas con ninguna media compresiva para personas que realizan vuelos largos

Pacientes o población: Cualquiera que realice un vuelo largo (que dure más de 6 horas)
Ámbitos: Vuelo aéreo internacional
Intervención: medias compresivas[1]
Comparación: Ninguna media

Resultados	Riesgos comparativos ilustrativos* (IC del 95%)		Efecto relativo (IC del 95%)	Número de participantes (estudios)	Calidad de las pruebas (GRADE)	Comentarios
	Riesgo asumido	Riesgo esperado				
	Ninguna media	Con medias				
Trombosis venosa profunda sintomática (TVP)	Ver Comentario	Ver Comentario	No estimable	2821 (9 estudios)	Ver Comentario	0 participantes desarrollaron TVP sintomática en estos estudios.
Trombosis venosa profunda asintomática	Población con bajo riesgo [2]		RR 0,10 (0,04 a 0,26)	2637 (9 estudios)	ÅÅÅÅ Alta	
	10 por 1000	1 por 1000 (0 a 3)				
	Población con alto riesgo [2]					
	30 por 1000	3 por 1000 (1 a 8)				
Trombosis venosa superficial	13 por 1000	6 por 1000 (2 a 15)	RR 0,45 (0,18 a 1,13)	1804 (8 estudios)	ÅÅÅ Moderada[3]	
Edema Valores posteriores al vuelo medidos en una escala de 0 (no edema) a 10 (edema máximo).	La puntuación media de edema varió entre los grupos control de 6 a 9.	La puntuación media de edema varió en los grupos intervención fue como promedio 4,7 menor (IC del 95%: -4,9 a -4,5).		1246 (6 estudios)	ÅÅ Baja[4]	
Embolismo pulmonar	Ver Comenta	Ver Comentari	No estimable	2821 (9	Ver Coment	0 participantes desarrollaron

	rio	o		estudios)	ario	embolismo pulmonar en estos estudios. [5]
Muerte	Ver Comentario	Ver Comentario	No estimable	2821 (9 estudios)	Ver Comentario	0 participantes murieron en estos estudios.
Efectos adversos	Ver Comentario	Ver Comentario	No estimable	1182 (4 estudios)	Ver Comentario	La tolerabilidad de las medias se describió como muy buena, sin quejas sobre los efectos secundarios en 4 estudios. [6]

* Las bases para el riesgo asumido se proporcionan en las notas al pie. El riesgo esperado (y su intervalo de confianza del 95%) se basa en el riesgo asumido en el grupo intervención y el efecto relativo de la intervención (y su IC del 95%).
IC: intervalo de confianza; RR: cociente de riesgos (risk ratio); GRADE: grados de las pruebas del Grupo de Trabajo GRADE (ver explicaciones).

[1] 1 Todas las medias utilizadas en los 9 estudios incluidos en esta revisión consistieron en medias compresivas por debajo de la rodilla. En cuatro ensayos la fuerza de compresión fue de 20-30 mmHg en el tobillo y de 10-20 mmHg en los otros cuatro ensayos. Las medias vienen en diferentes tamaños. Si una media está muy ajustada alrededor de la rodilla puede impedir el retorno venoso y provocar que la sangre se acumule alrededor de la rodilla. Las medias compresivas deben ajustarse de manera adecuada. Una media demasiado ajustada pudiera penetrar la piel durante un vuelo largo y provocar potencialmente ulceración y aumento del riesgo de TVP. Algunas medias pueden ser discretamente más gruesas que el recubrimiento normal de la pierna y pueden ser potencialmente limitantes cuando se utilizan zapatos ajustados. Es una buena idea usar las medias en la casa antes del viaje para asegurar un ajuste adecuado y cómodo. Las medias se colocaron 2 a 3 horas antes del vuelo en la mayoría de los ensayos. La disponibilidad y el costo de las medias pueden variar.

[2] Dos ensayos reclutaron participantes con alto riesgo, definidos como los que habían presentado episodios previos de TVP, trastornos de la coagulación, obesidad grave, movilidad limitada debido a problemas óseos o articulares, enfermedades malignas en los dos años previos, venas varicosas grandes o, en uno de los estudios, los participantes medían más de 190 cm y pesaban más de 90 kg. La incidencia para los 7 ensayos que excluyeron participantes con alto riesgo fue de 1,45% y la incidencia para los 2 ensayos que reclutaron participantes con alto riesgo (con al menos un factor de riesgo) fue de 2,43%. Estas cifras se redondearon a 10 y 30 por 1000, respectivamente.

[3] El intervalo de confianza cruza el valor que corresponde a la falta de diferencia y no descarta un aumento pequeño.

[4] La medición del edema no se validó ni cegó en la intervención. Todos estos estudios los realizaron los mismos investigadores.

[5] De haber pocos o ningún evento y si el número de participantes es grande, las valoraciones acerca de la calidad de las pruebas (particularmente las consideraciones acerca de la precisión) se pueden basar en el efecto absoluto. Aquí la calificación de la calidad se puede considerar "alta" si el resultado se evaluó de manera apropiada y el evento, de hecho, no ocurrió en ninguno de los 2821 participantes estudiados.

[6] Ninguno de los otros ensayos informó efectos adversos, aparte de los 4 casos de trombosis venosa superficial en las venas varicosas de la región de la rodilla comprimida por el borde superior de la media en un ensayo.

11.5.4 Producción de las tablas "Resumen de los resultados"

Un programa informático adicional, GRADEprofiler (GRADEpro), está disponible para ayudar a los revisores a preparar las tablas "Resumen de los resultados". GRDEpro puede recuperar datos de RevMan y combinarlos con los riesgos del grupo control ingresados por el usuario para producir los efectos relativos y los riesgos absolutos asociados con las intervenciones. Además, guía al usuario a través del proceso de una evaluación GRADE (ver el fichero de ayuda específica en GRADEpro) y produce una tabla que se puede importar fácilmente hacia RevMan como una tabla "Resumen de los resultados". La tabla se importa como una tabla especial (ver Sección 11.6), y no es posible modificarla en RevMan. Los revisores pueden crear alternativamente su propia tabla en RevMan.

11.5.5 Consideraciones estadísticas en las tablas "Resumen de los resultados"

Aquí se describe cómo se obtienen las medidas del efecto absolutas y relativas para los resultados dicotómicos. Los cocientes de riesgos, los odds ratios y las diferencias de riesgos son formas diferentes de comparar dos grupos con datos de resultados dicotómicos (ver Capítulo 9, Sección 9.2.2). Además, existen dos tipos diferentes de cocientes de riesgos, según qué evento (p.ej. "sí" o "no") es el centro del análisis (ver Capítulo 9, Sección 9.2.2.5). En presencia de un efecto de la intervención diferente de cero, si existe variación en los riesgos de los grupos control entre los estudios, no es posible que más de una de estas medidas sea con certeza la misma en cada estudio. En epidemiología siempre es de esperar que las medidas relativas del efecto sean más consistentes que las medidas absolutas del efecto de un escenario a otro. Actualmente hay pruebas empíricas que apoyan esta suposición (Engels 2000, Deeks 2001). Por este motivo, los metanálisis deben utilizar generalmente el cociente de riesgo o el odds ratio como la medida del efecto (ver Capítulo 9, Sección 9.4.4.4). De forma correspondiente, es probable que una estimación única de un efecto relativo sea un resumen más adecuado que una estimación única del efecto absoluto. Si un efecto relativo es en realidad consistente entre los estudios, entonces los diferentes riesgos de los grupos control tendrán diferentes implicaciones para el beneficio absoluto. Por ejemplo, si el cociente de riesgo es consistentemente 0,75 el tratamiento reduciría el riesgo del 80% en el grupo control al 60% en el grupo intervención (una reducción absoluta de 20 puntos porcentuales), pero reduciría el riesgo del 20% en el grupo control al 15% en el grupo intervención (una reducción absoluta de 5 puntos porcentuales).

Las tablas "Resumen de los resultados" se construyen alrededor de la suposición de un efecto relativo consistente. Por lo tanto, es importante considerar las implicaciones de este efecto para diferentes riesgos del grupo control. Para cualquier riesgo asumido en el grupo control es posible realizar la estimación de un riesgo esperado en el grupo control a partir del cociente de riesgos o del odds ratio metanalíticos. Es de señalar que los números proporcionados en la columna "Riesgo esperado" son específicos para los "Riesgos supuestos" en la columna adyacente.

Para el cociente de riesgos (CR) metanalítico y el riesgo control supuesto (RCS), el riesgo esperado de la intervención se obtiene de la siguiente manera:

Riesgo esperado de la intervención, por 1000 = 1000 × RCS× CR.

Como ejemplo, en la Figura 11.3a, el cociente de riesgos metanalítico es CR = 0,10 (IC del 95%: 0,04 a 0,26).
Si se supone que un riesgo control del RCS = 10 por 1000 = 0,01 se obtiene:

Riesgo esperado de la intervención, por 1000 = 1000 × 0,01 × 0,10 = 1,
como se indica en la Figura 11.5.a.

Para el odds ratio (OR) metanalítico y el riesgo control supuesto (RCS), el riesgo esperado de la intervención se obtiene de la siguiente manera:

$$\text{Corresponding intervention risk, per 1000} = 1000 \times \left(\frac{\text{OR} \times \text{ACR}}{1 - \text{ACR} + (\text{OR} \times \text{ACR})} \right).$$

Los límites de confianza superior e inferior para el riesgo esperado de la intervención se obtienen al reemplazar el CR o el OR por sus límites de confianza superior e inferior, respectivamente (p.ej. al reemplazar 0,10 por 0,04 y luego por 0,26 en el ejemplo anterior). Estos intervalos de confianza no incorporan la incertidumbre en los riesgos control supuestos.

Cuando se trata con los cocientes de riesgo, es fundamental que se utilice la misma definición de "evento" que se utilizó para el metanálisis. Por ejemplo, si el metanálisis se centró en "supervivencia" en lugar de "muerte" como evento, entonces los riesgos supuestos y correspondientes de la tabla "Resumen de los resultados" también se deben referir a "supervivencia".

En algunas circunstancias (poco frecuentes) en las cuales hay un motivo claro para suponer una diferencia de riesgo consistente en el metanálisis, en principio es posible presentarla para los "riesgos supuestos" relevantes y sus riesgos correspondientes, y presentar los efectos relativos correspondientes (diferentes) para cada riesgo asumido.

11.5.6 Contenidos detallados de una tabla "Resumen de los resultados"

11.5.6.1 Título y encabezado de la tabla

El título de cada tabla "Resumen de los resultados" debe especificar la pregunta clínica, enmarcada en términos de la población y que aclare exactamente qué comparación de intervenciones se realizan. En la Figura 11.5.a, la población son las personas que realizan vuelos muy largos, la intervención es el uso de medias compresivas y el control es ninguna media compresiva.

La primera fila de cada tabla "Resumen de los resultados" debe proporcionar la siguiente información en el "encabezamiento":

Pacientes o población: aclara de forma adicional la población (y posiblemente las subpoblaciones) de interés y de forma ideal la magnitud del riesgo de presentar el resultado adverso más importante al cual se dirige el tratamiento. Por ejemplo, los pacientes en un vuelo con un recorrido largo pueden tener riesgos diferentes de TVP, o los pacientes que utilicen ISRS pueden tener un riesgo diferente de efectos secundarios, o los pacientes con fibrilación auricular pueden tener un riesgo anual bajo (< 1%), moderado (1% a 4%) o alto (> 4%) de accidente cerebrovascular.

Ámbito: debe especificar cualquier característica particular del ámbito en el cual se realizaron los estudios que pudiera limitar la aplicabilidad del resumen de los resultados a otros ámbitos; p.ej. la atención primaria en Europa y Norteamérica.

Intervención: la intervención experimental.

Comparación: la intervención (comparación) control (incluido ningún tratamiento específico).

11.5.6.2 Resultados

Las filas de una tabla "Resumen de los resultados" debe incluir todos los resultados deseables y no deseables (enumerados en orden de importancia), que son fundamentales para la toma de decisiones, hasta un máximo de siete resultados. Si hay un número excesivo de resultados en la revisión, los revisores deberán omitir los resultados menos importantes. Se deben proporcionar detalles de las escalas y los puntos temporales en los que se realizan las medidas. Los revisores deben haber decidido qué resultados son importantes para la tabla "Resumen de los resultados" durante el desarrollo del protocolo y antes de realizar la revisión. Sin embargo, los revisores deben estar alerta ante la posibilidad de que la importancia de un resultado (p.ej. un efecto adverso grave) se conozca sólo después de haber redactado el protocolo o haber realizado el análisis, y deberían realizar las acciones apropiadas para incluirlos en la tabla "Resumen de los resultados". Es

de señalar que los revisores deben enumerar estos resultados en la tabla *estén disponibles los datos o no*.

Se deben incluir los eventos adversos graves, pero es posible combinar los eventos adversos menores y describirlos en una nota al pie (cabe señalar que no es apropiado colocar los eventos juntos a menos que se conozca que son independientes). Los puntos temporales múltiples pueden ser un problema particular. En general, para que la tabla sea sencilla, sólo se deben presentar en puntos temporales múltiples los resultados que sean fundamentales para la toma de decisiones. Los restantes se deben presentar en un punto temporal común.

Las medidas de resultado continuas se pueden mostrar en la tabla "Resumen de los resultados"; los revisores deben esforzarse para hacer que las mismas sean interpretables para la audiencia (ver Capítulo 12, Sección 12.6). Lo anterior requiere que las unidades sean claras y fácilmente interpretables, por ejemplo, días de dolor o frecuencia de cefalea. Sin embargo, muchos instrumentos de medición no son fácilmente interpretables por los médicos no especializados o los pacientes, por ejemplo, las puntuaciones del Inventario de Depresión de Beck o la puntuación de la calidad de vida. Para este tipo de medidas una presentación más fácil de interpretar pudiera incluir la conversión de un resultado continuo en dicotómico, como > 50% de mejoría (ver Capítulo 12, Sección 12.6).

11.5.6.3 Riesgos comparativos ilustrativos 1: Riesgo asumido (con la intervención control)

Los revisores deben proporcionar hasta tres riesgos típicos para los participantes que reciben la intervención control. Se recomienda que los mismos se presenten como el número de personas que experimentan el evento por 1000 personas (frecuencia natural). Se puede utilizar una alternativa mayor de 1000 para los eventos poco frecuentes, o también es posible utilizar 100 para eventos más frecuentes. Los riesgos supuestos de la intervención control se pueden basar en las evaluaciones de los riesgos típicos en diferentes grupos de pacientes o en diferentes duraciones del seguimiento. De manera ideal los riesgos reflejarían los grupos que los médicos pueden identificar con facilidad según las características que presentan. Una nota al pie debe especificar la fuente o la justificación para cada riesgo del grupo control, incluido el período de tiempo que corresponda, de ser apropiado. En la Figura 11.5.a los médicos pueden diferenciar fácilmente los individuos con factores de riesgo para la trombosis venosa profunda de los que no presentan dicho riesgo. Si se conoce que existe poca variación en el riesgo basal los revisores pueden utilizar la mediana del riesgo del grupo control entre los estudios.

11.5.6.4 Riesgos comparativos ilustrativos 2: Riesgo esperado (con la intervención experimental)

Para los resultados dicotómicos, se debe proporcionar un riesgo absoluto correspondiente para cada riesgo asumido en la columna precedente, junto con un intervalo de confianza. Este riesgo absoluto con la intervención (experimental) habitualmente se derivará del resultado del metanálisis presentado en la columna del efecto relativo (ver Sección 11.5.6.5). Las fórmulas se proporcionan en la Sección 11.5.5. Los revisores deben presentar el efecto absoluto en el mismo formato que el de los riesgos supuestos con la intervención control (ver Sección 11.5.6.3), p.ej. como el número de personas que presentaron el evento por 1000 personas.

Para los resultados continuos se debe presentar una diferencia de medias o una diferencia de medias estandarizada con su intervalo de confianza. Éstas habitualmente se obtienen directamente de un metanálisis. Se debe utilizar un texto explicativo para aclarar el significado, como en la Figura 11.5.a.

11.5.6.5 Efecto relativo (IC del 95%)

El efecto relativo será habitualmente un cociente de riesgos o un odds ratio (ocasionalmente un cociente de riesgos instantáneo), con su intervalo de confianza del 95% acompañante,

obtenido de un metanálisis realizado según la misma medida del efecto. Los cocientes de riesgos y los odds ratios son similares cuando los riesgos de la intervención control son bajos y los efectos son pequeños, pero difieren considerablemente cuando los mismos aumentan. El metanálisis puede incluir una suposición de los efectos fijos o aleatorios, según lo que los revisores consideren apropiado.

11.5.6.6 Número de participantes (estudios)

Esta columna debe incluir el número de participantes evaluados en los estudios incluidos para cada resultado y el número de estudios correspondientes que contribuyeron con estos participantes.

11.5.6.7 Calidad de las pruebas (GRADE)

Los revisores comentarán acerca de la calidad del cuerpo de las pruebas como "Alta", "Moderada", "Baja" o "Muy baja". Lo anterior se basa en la valoración, pero el proceso de valoración opera con una estructura transparente y se describe en el Capítulo 12 (Sección 12.2). Como ejemplo, la calidad sería "Alta" si el resumen proviene de varios ensayos aleatorios con bajo riesgo de sesgo, pero la calificación de la calidad es menor si hay dudas acerca del diseño o la implementación, imprecisión, inconsistencias, falta de direccionalidad o sesgo de reporte selectivo. Los revisores deben utilizar el sistema específico para la calificación de las pruebas desarrollado por la colaboración GRADE (GRADE Working Group 2004) que se describe en detalle en el Capítulo 12 (sección 12.2). Otras valoraciones diferentes de la calidad "Alta" se deben aclarar adecuadamente mediante notas al pie o en la columna de Comentarios de la tabla "Resumen de los resultados" (ver Figura 11.5.a).

11.5.6.8 Comentarios

El objetivo del campo Comentarios es proporcionar comentarios adicionales que ayuden a interpretar la información o los datos identificados en la fila. Por ejemplo, pueden ser sobre la validez de la medida de resultado o la presencia de variables asociadas con la magnitud del efecto. Aquí se deben hacer constar las advertencias importantes acerca de los resultados. No todas las filas necesitarán comentarios, por lo que es mejor dejarlas en blanco si no se justifica un comentario.

11.6 Tablas adicionales

La sección Tablas adicionales proporciona una forma flexible de crear tablas, lo que permite la presentación de resultados de los ensayos y los metanálisis, así como otras investigaciones metanalíticas (como los análisis de metarregresión). Los resultados importantes de todas las Tablas adicionales se deben resumir en la sección Resultados del texto de la revisión.

11.7 Presentación de los resultados en el texto

11.7.1 Resultados de los metanálisis

La sección Resultados se debe organizar para seguir el orden de las comparaciones y los resultados especificados en el protocolo de manera que aborde explícitamente los objetivos de la revisión. El texto debe presentar los resultados generales de una manera lógica y sistemática: no se debe basar demasiado en las tablas o figuras, ni referirse constantemente a ellas para dar una imagen clara de los resultados de la revisión. En su lugar las tablas se deben utilizar como un recurso adicional que pudiera proporcionar detalles adicionales. Sin embargo, se debe evitar la repetición excesiva en el texto de los datos que también se proporcionan en las tablas o figuras.

No se debe hacer un énfasis exagerado en las respuestas a los análisis *post hoc* y a preguntas menos importantes para las cuales puede existir una gran cantidad de datos. Los análisis *post hoc* se deben identificar siempre como tales. Los revisores deben dejar claro en

la sección Resultados el método de análisis utilizado parta cada resultado citado (en particular, la selección de la medida del efecto, la dirección de un efecto beneficioso y el modelo de metanálisis utilizado), aunque los propios métodos analíticos se deben describir en la sección Métodos. Los resultados siempre se deben acompañar por una medida de incertidumbre como el intervalo de confianza del 95%. El resumen debe condensar los resultados solamente de las comparaciones y resultados más importantes, y no informar selectivamente los que presenten los resultados más significativos. También es útil indicar la cantidad de información (número de estudios y participantes) en los cuales se basaron los análisis.

Cada figura y Tabla adicional se debe citar explícitamente en el texto. Cuando se haga referencia a los resultados en una figura, tabla o diagrama de bosque de "Datos y análisis" que no se ha seleccionado como figura, la figura, tabla o análisis se debe citar en el texto.

Los revisores deben considerar la presentación de los resultados en un formato que sea fácil de interpretar. Por ejemplo, los odds ratios y las diferencias de medias estandarizadas por sí mismas no son medidas adecuadas para la aplicación directa en la práctica clínica, pero se pueden expresar nuevamente de una forma más accesible. Ver Capítulo 12 (Secciones 12.5 y 12.6)

11.7.2 Resultados sin metanálisis

Los métodos para el metanálisis permiten la cuantificación de la dirección del efecto, el tamaño y la consistencia del efecto (ver Capítulo 9, Sección 9.1). Si no hay datos numéricos adecuados disponibles para el metanálisis o si los metanálisis se consideran inapropiados, a menudo todavía es posible examinar estos dominios para proporcionar una evaluación sistemática de las pruebas disponibles.

Una evaluación narrativa de las pruebas puede ser un reto, especialmente si la revisión incluye un gran número de estudios; si los estudios por sí mismos examinan intervenciones y resultados complejos; o si existe una gran variación en los efectos de la intervención. Por lo tanto, es posible que los patrones del efecto y las semejanzas o diferencias entre los estudios no sean evidentes de inmediato. La adopción de un enfoque sistemático para la presentación es importante para que los resultados de una revisión tengan sentido. Si se proporciona un párrafo descriptivo para los resultados de cada estudio, éste debe ser consistente, incluir los mismos elementos de información para cada estudio y presentarlos en el mismo orden. Se recomienda la organización de los estudios en grupos o conjuntos (p.ej. por tipo de intervención, grupos de población, ámbitos, etc.) si se ha incluido un gran número de estudios (p.ej. más de 20) en la revisión, lo que puede hacer más fácil el proceso de descripción narrativa de los resultados. También puede permitir la identificación de patrones en los resultados, dentro o entre los grupos formados.

11.8 Redacción de un resumen

Todas las revisiones completas deben incluir un resumen de no más de 400 palabras. El resumen debe ser tan breve como sea posible sin sacrificar contenido importante. Los resúmenes de las revisiones Cochrane se publican en MEDLINE y el Science Citation Index, y están libremente disponibles en Internet. Por lo tanto, es importante que se puedan leer como documentos individuales.

El resumen debe hacer un compendio de los métodos, los resultados y las conclusiones clave de la revisión y no debe contener información que no esté en la revisión. Los vínculos con otras partes de la revisión (como las referencias, los estudios, las tablas y las figuras) no se deben incluir en el resumen. Un ejemplo hipotético de un resumen se muestra en el Recuadro 11.8.a.

Los resúmenes deben estar dirigidos principalmente a quienes toman decisiones en la atención sanitaria (médicos, consumidores informados y los responsables de políticas de salud), en lugar de sólo a los investigadores. La terminología debe ser razonablemente comprensible para un auditorio general, más que especializado, en la atención sanitaria. Se deben evitar las abreviaturas, excepto cuando sean ampliamente conocidas (p.ej. VIH). Cuando sea fundamental, se deben explicar las abreviaturas (con las siglas entre paréntesis) cuando se utilicen por primera vez. Siempre que sea posible se deben utilizar los nombres de los fármacos y las intervenciones que se comprendan a nivel internacional. No se deben utilizar nombres comerciales.

El contenido bajo cada encabezado del resumen debe ser el siguiente:

Antecedentes: debe ser de una o dos oraciones para explicar el contexto o elaborar el objetivo y la justificación de la revisión. Si esta versión de la revisión es una actualización de una anterior es útil incluir una oración como "Ésta es una actualización de una revisión Cochrane publicada en AÑO, y actualizada previamente en AÑO".

Objetivos: debe ser un planteamiento preciso del objetivo primario de la revisión, idealmente en una única oración que vincule los Objetivos con el texto principal de la revisión. De ser posible, el estilo debe ser en forma de "Evaluar los efectos de *[intervención o comparación]* para *[problema de salud]* para/en *[tipo de personas, enfermedad o problema y ámbito si se especifica]*".

Métodos de búsqueda: debe enumerar las fuentes y las fechas de la última búsqueda, para cada fuente, mediante la forma activa "Se buscó…" o si fue solamente un revisor, se puede utilizar la forma pasiva, por ejemplo, "Las bases de datos X, Y, Z fueron revisadas". Los términos de búsqueda no se deben enumerar aquí. Si se utilizó el Registro Especializado del GCR, el mismo se debe enumerar primero como "Registro Especializado del Grupo Cochrane X". El orden para enumerar otras bases de datos debe ser Registro Cochrane Central de Ensayos Controlados, MEDLINE, EMBASE, otras bases de datos. Se debe proporcionar el rango de las fechas de la búsqueda para cada base de datos. Para el Registro Cochrane Central de Ensayos Controlados debe ser como "Registro Cochrane Central de Ensayos Controlados (*The Cochrane Library* 2007, Número 1)". Para la mayoría de las otras bases de datos como MEDLINE debe ser como "MEDLINE (enero de 1966 a diciembre de 2006)". La búsqueda en listas de bibliografía para identificar citas relevantes se puede cubrir en una frase genérica como "listas de referencias de los artículos". Si hubo alguna limitación según el idioma o el estado de publicación, la misma se debe enumerar. Si se estableció contacto con individuos u organizaciones para localizar estudios, se debe señalar y es preferible utilizar "Se estableció contacto con compañías farmacéuticas" en lugar de enumerar todas las compañías farmacéuticas con las que se estableció contacto. Se debe señalar si se realizaron específicamente búsquedas manuales en las revistas para la revisión, pero no se debe enumerar la búsqueda manual que ayuda a conformar el Registro Especializado del GCR.

Criterios de selección: se debe proporcionar como "*[tipo de estudios]* o *[tipo de intervención o comparación]* en *[enfermedad, problema o tipo de personas]*". Los resultados se deben enumerar aquí solamente si la revisión se limitó a resultados específicos.

Obtención y análisis de los datos: se debe limitar a cómo se extrajeron y evaluaron los datos y no incluir detalles de qué datos se extrajeron. Esta sección debe cubrir si la extracción de los datos y la evaluación del riesgo de sesgo las realizó más de una persona. Se debe señalar aquí si los revisores establecieron contacto con los investigadores para obtener información adicional, así como qué pasos se dieron, si hubo alguno, para identificar los efectos adversos.

Resultados principales: Esta sección debe comenzar con el número total de estudios y participantes incluidos en la revisión, y con detalles breves pertinentes para la

interpretación de los resultados (por ejemplo, el riesgo general de sesgo en los estudios o un comentario sobre la comparabilidad de los estudios, de ser apropiado). Se debe abordar el objetivo primario y limitarse a los resultados cualitativos y cuantitativos principales (generalmente se incluyen no más de seis resultados clave). Los resultados incluidos se deben seleccionar sobre la base de cuál es más probable que ayude a tomar una decisión sobre si utilizar o no una intervención particular. Los efectos adversos se deben incluir si se analizan en la revisión. De ser necesario, se debe señalar el número de estudios y participantes que contribuyen con los resultados por separado, junto con las inquietudes sobre la calidad de las pruebas específicas para estos resultados. Los resultados se deben expresar de forma narrativa, así como cuantitativamente si los resultados numéricos no son claros ni intuitivos (como los de los análisis de las diferencias de medias estandarizadas). Los estadísticos resumen descritos en esta sección deben ser los mismos que los seleccionados por defecto para la revisión, y se deben presentar de una manera estándar, como "odds ratio 2,31 (intervalo de confianza del 95%: 1,13 a 3,45)". Idealmente, se deben informar los riesgos de presentar los eventos (porcentaje) o los promedios (para los datos continuos) para ambos grupos de comparación. Si en la revisión no se calcularon los resultados globales, se puede proporcionar una evaluación o una descripción cualitativa del rango de valores y el patrón de los resultados. Sin embargo, se debe evitar el "recuento de votos" en el cual se informa el número de estudios "positivos" y "negativos".

Conclusiones de los revisores: el objetivo principal de la revisión debe ser presentar la información, en lugar de ofrecer consejo o recomendaciones. Las Conclusiones de los revisores deben ser sucintas y provenir directamente de los resultados de la revisión, de manera que reflejen directa y evidentemente los resultados principales. En general no se deben hacer suposiciones acerca de las circunstancias prácticas, valores, preferencias, compensaciones, y se deben evitar los consejos o recomendaciones. Se debe señalar cualquier limitación importante de los datos o los análisis. Se deben incluir las conclusiones importantes acerca de las implicaciones para la investigación, si las mismas no son evidentes.

Recuadro 11.8.a: Ejemplo hipotético de un resumen
(Para la revisión "A versus B para el tratamiento de la gripe en adultos" por Peach A, Apricot D, Plum P.)

Antecedentes
A y B tienen propiedades antivirales, pero no se utilizan ampliamente debido al conocimiento incompleto de sus propiedades y a inquietudes acerca de los posibles efectos adversos. Ésta es una actualización de una revisión Cochrane publicada por primera vez en 1999 y actualizada previamente en 2006.
Objetivos
Evaluar los efectos de A y B en adultos con gripe.
Métodos de búsqueda
Se buscó en el Registro Especializado del Grupo Cochrane de Infecciones Respiratorias Agudas (15 de febrero 2007), el Registro Cochrane Central de Ensayos Controlados (*The Cochrane Library* Número 1, 2007), MEDLINE (enero de 1966 a enero de 2007), EMBASE (enero de 1985 a diciembre de 2006) y las listas de referencias de los artículos. También se estableció contacto con los fabricantes e investigadores en el campo.
Criterios de selección
Estudios aleatorios y cuasialeatorios que comparen A o B con placebo, o comparen dosis o esquemas de A o B en adultos con gripe.
Obtención de los datos
Dos revisores evaluaron de forma independiente la calidad del ensayo y extrajeron los datos. Se estableció contacto con los autores de los estudios para solicitar información adicional. Se obtuvo información de los ensayos sobre los efectos adversos.
Resultados principales

Se incluyeron 17 ensayos donde participaron 689 personas. Cinco ensayos que reclutaron 234 personas compararon A con placebo. Comparado con placebo, A acortó significativamente la duración de la fiebre en el 23% (en 1,0 días; intervalo de confianza del 95%: 0,73 a 1,29). Seis ensayos donde participaron 256 personas compararon B con placebo. B acortó significativamente la duración de la fiebre en el 33% comparado con placebo (en 1,27 días; intervalo de confianza del 95%: 0,77 a 1,77). La escasa cantidad de información disponible para comparar directamente A y B (dos ensayos con 53 personas) no indicó diferencias en la eficacia de ambos fármacos, aunque los intervalos de confianza fueron muy amplios. Según cuatro ensayos de 73 personas, los efectos en el sistema nervioso central fueron significativamente más frecuentes con A que con B (riesgo relativo 2,58; intervalo de confianza del 95%: 1,54 a 4,33).
Conclusiones de los revisores
A y B parecen ser efectivos en el tratamiento de la gripe. No hay pruebas suficientes para determinar si uno es más efectivo que el otro. Ambos fármacos parecen ser relativamente bien tolerados, aunque B parece ser más seguro.

11.9 Redacción de un resumen en lenguaje sencillo

11.9.1 Acerca de los resúmenes en lenguaje sencillo

El resumen en lenguaje sencillo tiene como objetivo resumir la revisión en un estilo directo que los consumidores puedan comprender. Los resúmenes en lenguaje sencillo están libremente disponibles en Internet, por lo que a menudo se consultarán como documentos individuales. Los resúmenes en lenguaje sencillo tienen dos partes: un título y un cuerpo de texto.

Generalmente el primer borrador del resumen en lenguaje sencillo lo deben redactar los revisores y enviarlos con la revisión al GCR pertinente. Este borrador puede estar sujeto a cambios y los revisores deben prever una o más repeticiones. Muchos GCR tienen herramientas para la redacción del resumen en lenguaje sencillo dentro de su equipo editorial. Cuando lo anterior no esté disponible, hay un servicio de apoyo central para ayudar a los GCR en la redacción y la edición. Este servicio está coordinado por la Red de Consumidores Cochrane, pero los revisores que necesiten ayuda para redactar un resumen en lenguaje sencillo deben establecer contacto con su GCR.

Existe información adicional disponible sobre el proceso de realización de los resúmenes en lenguaje sencillo en el *Manual Cochrane* (disponible en www.cochrane.org/admin/manual.htm).

11.9.2 Título en lenguaje sencillo

La primera parte de un resumen en lenguaje sencillo es un replanteamiento del título de la revisión donde se utilicen términos en lenguaje sencillo. Lo anterior debe incluir los participantes y la intervención (y el resultado, cuando se incluye en el título de la revisión). Por ejemplo, el título de revisión "Fármacos anticolinérgicos versus otras medicaciones para el síndrome de vejiga hiperactiva en adultos" pudiera cambiarse en lenguaje sencillo a "Fármacos para el síndrome de vejiga hiperactiva". Cuando el título de la revisión sea fácilmente comprensible, sencillamente se debe replantear como título en lenguaje sencillo, p. ej. "Intervenciones para reducir el daño del uso continuado del tabaco".

El título en lenguaje sencillo no debe ser declarativo (no debe reflejar las conclusiones de la revisión). Se debe redactar tipo oración (es decir, con una mayúscula al inicio del título y para los nombres, pero el resto en letra minúscula; ver ejemplos anteriormente), no debe tener más de 256 caracteres y no debe terminar con una interrupción brusca.

11.9.3 Texto resumen

La segunda parte o cuerpo del resumen en lenguaje sencillo no debe tener más de 400 palabras y debe incluir:

- Un planteamiento acerca de por qué la revisión es importante: por ejemplo, la definición y los antecedentes del problema sanitario, los signos y síntomas, la prevalencia, la descripción de la intervención y la justificación para su uso.
- Los resultados principales de la revisión: puede incluir resúmenes numéricos cuando la revisión ha informado resultados en forma numérica, pero se deben presentar en un formato general y fácilmente comprensible. Los resultados que se presentan en el resumen en lenguaje sencillo no deben ser diferentes de los presentados en la revisión (es decir, no deben aparecer resultados nuevos en el resumen). De ser posible se debe proporcionar una indicación del número de ensayos y participantes en los cuales se basan los resultados.
- Un comentario de cualquier efecto adverso.
- Un comentario breve de cualquier limitación de la revisión (por ejemplo, ensayos en poblaciones muy específicas o métodos deficientes de los ensayos incluidos).

Al final del resumen en lenguaje sencillo los revisores deben proporcionar vínculos web (por ejemplo, a otras ayudas para información o decisión en sitios web del GCR, siempre que cumplan con la política de la Colaboración Cochrane para los vínculos web. Los gráficos o cuadros no se deben incluir en los resúmenes en lenguaje sencillo. Al igual que con cualquier otro componente de una revisión Cochrane, los resúmenes en lenguaje sencillo deben seguir el formato de la Guía de Estilo Cochrane (disponible en www.cochrane.org/style).

11.10 Información del capítulo

Autores: Holger J Schünemann, Andrew D Oxman, Julian PT Higgins, Gunn E Vist, Paul Glasziou y Gordon H Guyatt del Grupo Cochrane de Métodos de Aplicabilidad y Recomendaciones y del Grupo Cochrane de Métodos Estadísticos.

La versión en inglés de este capítulo se debe citar como: Schünemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Chapter 11: Presenting results and 'Summary of findings' tables. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: El profesor Penny Hawe contribuyó al texto sobre los efectos adversos. Jon Deeks aportó contribuciones útiles. Sally Green, Janet Wale y Gill Gyte desarrollaron las guías sobre los resúmenes en lenguaje sencillo y también utilizaron las guías para la síntesis narrativa de Rebecca Ryan y el Grupo de Revisión de Comunicación y Consumidores. El material sobre la redacción de los resúmenes se basa en las versiones anteriores del *Manual*. Para detalles de los autores y editores anteriores refiérase a la Sección 1.4. Para detalles del Grupo Cochrane de Métodos de Aplicabilidad y Recomendaciones ver Capítulo 12 (Cuadro 12.8.a); para el Grupo Cochrane de Métodos Estadísticos ver Capítulo 9 (Cuadro 9.8.a).

Conflicto de interés: Holger Schünemann, Andrew Oxman, Gunn Vist, Paul Glasziou y Gordon Guyatt han desempeñado, en diferentes grados, funciones directivas en el Grupo de Trabajo GRADE, del cual han surgido muchas de las ideas de las tablas "Resumen de los resultados".

11.11 Referencias

Clarke 2006
Clarke M, Hopewell S, Juszczak E, Eisinga A, Kjeldstrøm M. Compression stockings for preventing deep vein thrombosis in airline passengers. *Cochrane Database of Systematic Reviews* 2006, Issue 2. Art No: CD004002.

Deeks 2001
Deeks JJ, Altman DG. Effect measures for meta-analysis of trials with binary outcomes. In: Egger M, Davey Smith G, Altman DG (editors). *Systematic Reviews in Health Care: Meta-analysis in Context* (2nd edition). London (UK): BMJ Publication Group, 2001.

Engels 2000
Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Statistics in Medicine* 2000; 19: 1707-1728.

GRADE Working Group 2004
GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328: 1490-1494.

Hind 2007
Hind D, Booth A. Do health technology assessments comply with QUOROM diagram guidance? An empirical study. BMC Med Res Methodol 2007; 7: 49.

Lewis 2001
Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ* 2001; 322: 1479-1480.

Liberati 2009
Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Medicine 2009; 6: e1000100.

ÍNDICE

CAPÍTULO 12: INTERPRETACIÓN DE LOS RESULTADOS Y OBTENCIÓN DE LAS CONCLUSIONES

PUNTOS CLAVE

12.1 INTRODUCCIÓN

12.2 EVALUACIÓN DE LA CALIDAD DE UN CONJUNTO DE PRUEBAS

12.2.1 El enfoque GRADE
Tabla 12.2.a: Niveles de calidad del conjunto de las pruebas en el enfoque GRADE
Tabla 12.2.b: Factores que pueden disminuir el nivel de calidad de un conjunto de pruebas
Tabla 12.2.c: Factores que pueden aumentar el nivel de calidad de un conjunto de pruebas
12.2.2 Factores que disminuyen el nivel de calidad de un conjunto de pruebas
Tabla 12.2.d: Guía adicional para el factor 1 (de 5) en una evaluación GRADE: de las evaluaciones del riesgo de sesgo a las valoraciones acerca de las limitaciones del estudio para los resultados principales
12.2.3 Factores que aumentan el nivel de calidad de un conjunto de pruebas

12.3 ASPECTOS DE LA APLICABILIDAD

12.3.1 La función del revisor
12.3.2 Variación biológica
12.3.3 Variación en el contexto y la cultura
12.3.4 Variación en el cumplimiento
12.3.5 Variación en los valores y preferencias

12.4 INTERPRETACIÓN DE LOS RESULTADOS DE LOS ANÁLISIS ESTADÍSTICOS

12.4.1 Intervalos de confianza
12.4.2 Valores de p y significación estadística

12.5 INTERPRETACIÓN DE LOS RESULTADOS DICOTÓMICOS (INCLUIDOS LOS NÚMEROS NECESARIOS A TRATAR)

12.5.1 Reducciones relativas y absolutas del riesgo
12.5.2 Más acerca del número necesario a tratar (NNT)
12.5.3 Expresión de las reducciones del riesgo absoluto
12.5.4 Cálculos
- 12.5.4.1 Cálculo del NNT a partir de la diferencia de riesgo (DR)
- 12.5.4.2 Cálculo de la reducción del riesgo absoluto o el NNT a partir de un cociente de riesgos (CR)
- 12.5.4.3 Cálculo de la reducción del riesgo absoluto o el NNT a partir de un odds ratio (OR)
- 12.5.4.4 Cálculo del cociente de riesgos a partir del odds ratio (OR)
- 12.5.4.5 Cálculo de los límites de confianza

12.6 INTERPRETACIÓN DE LOS RESULTADOS CONTINUOS (INCLUIDAS LAS DIFERENCIAS DE MEDIAS ESTANDARIZADAS)

12.6.1 Metanálisis con resultados continuos
12.6.2 Reexpresión de las DME mediante la regla empírica para los tamaños del efecto
12.6.3 Reexpresión de las DME mediante la transformación del odds ratio
Tabla 12.6.a: NNT equivalentes para DME específicas para varias “proporción de personas con mejoría” en el grupo control
12.6.4 Reexpresión de las DME mediante un instrumento familiar

12.7 CONCLUSIONES

12.7.1 Sección Conclusiones de una revisión Cochrane
12.7.2 Implicaciones para la práctica
12.7.3 Implicaciones para la investigación
12.7.4 Errores frecuentes al establecer las conclusiones

12.8 INFORMACIÓN DEL CAPÍTULO

Recuadro 12.8.a: El Grupo Cochrane de Métodos de Aplicabilidad y Recomendaciones

12.9 REFERENCIAS

CAPÍTULO 12: INTERPRETACIÓN DE LOS RESULTADOS Y OBTENCIÓN DE LAS CONCLUSIONES

Autores: Holger J Schünemann, Andrew D Oxman, Gunn E Vist, Julian PT Higgins, Jonathan J Deeks, Paul Glasziou y Gordon H Guyatt del Grupo Cochrane de Métodos de Aplicabilidad y Recomendaciones.

Puntos clave

- El enfoque GRADE, adoptado por la Colaboración Cochrane especifica cuatro niveles de calidad (alto, moderado, bajo y muy bajo), donde la calificación de calidad más alta corresponde a un conjunto de pruebas basado en ensayos clínicos aleatorios. Los revisores pueden disminuir la calificación de las pruebas de los ensayos clínicos aleatorios según la presencia de cinco factores y aumentar la calidad de las pruebas de los estudios observacionales según tres factores.
- Las calificaciones de la calidad se hacen de forma independiente para cada resultado.
- En este capítulo se describen los métodos para computar, presentar e interpretar los efectos relativos y absolutos para los datos dicotómicos, incluido el número necesario a tratar (NNT).
- Para las medidas de resultado continuas los revisores pueden presentar los resultados agrupados para los estudios mediante las mismas unidades, la diferencia de medias estandarizada y los tamaños del efecto cuando los estudios utilizan el mismo constructo pero diferentes escalas, así como los odds ratios después de la transformación de las diferencias de medias estandarizadas.
- Los revisores no deben describir resultados como “no estadísticamente significativos” o “no significativos”, sino informar los intervalos de confianza con el valor exacto de p.
- Los revisores no deben hacer recomendaciones, pero pueden (después de describir la calidad de las pruebas y el equilibrio entre los beneficios y los daños) señalar diferentes acciones que pudieran ser consistentes con patrones particulares de valores y preferencias.

12.1 Introducción

El objetivo de las revisiones Cochrane es facilitar la toma de decisiones en la atención sanitaria por parte de los pacientes, el público general, los médicos, administradores y responsables de políticas de salud. Un planteamiento claro de los hallazgos, una discusión meditada y una presentación clara de las conclusiones de los revisores son partes importantes de la revisión. En particular, los siguientes aspectos pueden ayudar a las personas a tomar decisiones mejor informadas y aumentar la utilidad de las revisiones Cochrane:

- la información de todos los resultados importantes, incluidos los efectos adversos;
- la calidad de las pruebas para cada uno de estos resultados, y cómo se aplica a poblaciones e intervenciones específicas; y
- la aclaración de la manera en la cual los valores y preferencias particulares pueden influir en el equilibrio de los beneficios y los daños, la carga y los costos de la intervención.

Una tabla de "Resumen de los hallazgos" descrita en el Capítulo 11 (Sección 11.5) proporciona piezas clave de información en un formato rápido y accesible. Se recomienda que los revisores incluyan dichas tablas en las revisiones Cochrane, y aseguren una descripción suficiente de los estudios y los metanálisis que apoyen el contenido de las mismas. La sección Discusión del texto debe proporcionar consideraciones complementarias. Los revisores deben utilizar cinco subencabezados para asegurar que incluyen suficientes aspectos en la sección Discusión y que colocan la revisión en un contexto apropiado. Dichos subencabezados son: "Resumen de los resultados principales (beneficios y daños)", "Compleción general y aplicabilidad de las pruebas", "Calidad de las pruebas", "Sesgos potenciales en el proceso de revisión" y "Acuerdos y desacuerdos con otros estudios y revisiones". Las conclusiones de los revisores se dividen en "Implicaciones para la práctica" e "Implicaciones para la investigación".

Debido a que las revisiones Cochrane tienen un auditorio internacional, la discusión y las conclusiones de los revisores deben, en lo posible, asumir una perspectiva internacional amplia y proporcionar una guía sobre cómo se pueden aplicar los resultados en diferentes ámbitos, en lugar de limitarse a circunstancias nacionales o locales específicas. Las diferencias culturales y económicas pueden tener una función importante al determinar el mejor curso de acción. Además, los individuos dentro de las sociedades tienen valores y preferencias que varían ampliamente con respecto a los estados de salud y al uso de los recursos de la sociedad para alcanzar determinados estados de salud. Incluso en el caso de tener los mismos valores y preferencias, las personas pueden interpretar las mismas pruebas de investigación de manera diferente. Por todos estos motivos, a menudo personas diferentes tomarán decisiones diferentes basadas en las mismas pruebas.

Por lo tanto, el objetivo de la revisión debe ser presentar la información y ayudar a la interpretación en lugar de ofrecer recomendaciones. La discusión y las conclusiones deben ayudar a las personas a comprender las implicaciones de las pruebas en relación con las decisiones prácticas y aplicar los resultados a su situación específica. Los revisores deben evitar recomendaciones específicas que dependan de suposiciones acerca de la disponibilidad de los recursos y los valores. Sin embargo, los revisores deben ayudar a la toma de decisiones mediante la exposición de diferentes escenarios que describan ciertas estructuras de valores.

En este capítulo se aborda primero uno de los aspectos clave de la interpretación de los hallazgos que también es fundamental para completar una tabla "Resumen de los hallazgos": la calidad de las pruebas relacionada con cada uno de los resultados. Luego se proporcionan consideraciones más detalladas de los aspectos alrededor de la aplicabilidad y la interpretación de los resultados numéricos y se brindan indicaciones para la presentación de las conclusiones de los revisores.

12.2 Evaluación de la calidad de un conjunto de pruebas

12.2.1 El enfoque GRADE

El Grades of Recommendation, Assessment, Development and Evaluation Working Group (GRADE Working Group) ha desarrollado un sistema para calificar la calidad de las pruebas (GRADE Working Group 2004, Schünemann 2006b, Guyatt 2008a, Guyatt 2008b). Cerca de 20 organizaciones incluida la Organización Mundial de la Salud (OMS), el American College of Physicians, el American College of Chest Physicians (ACCP), la American Endocrine Society, la American Thoracic Society (ATS), la Canadian Agency for Drugs and Technology in Health (CADTH), BMJ Clinical Evidence, los National Institutes of Health and Clinical Excellence (NICE) en el RU y UpToDate® han adoptado el sistema GRADE en su formato original o con modificaciones menores (Schünemann 2006b, Guyatt 2006a, Guyatt 2006b). El BMJ recomienda que los autores de guías clínicas utilicen el sistema GRADE (www.bmj.com/advice/sections.shtml). La Colaboración Cochrane ha adoptado los principios del sistema GRADE para la evaluación de la calidad de las pruebas para los resultados informados en las revisiones sistemáticas. Esta evaluación está en proceso de introducción conjuntamente con la tabla “Resumen de los hallazgos” (ver Capítulo 11, Sección 11.5).

Para los objetivos de las revisiones sistemáticas, el enfoque GRADE define la calidad del conjunto de las pruebas como el grado en el cual es posible confiar en que una estimación del efecto o asociación están cerca de la magnitud específica de interés. La calidad del conjunto de las pruebas incluye la consideración del riesgo de sesgo dentro del estudio (calidad metodológica), la direccionalidad de las pruebas, la heterogeneidad, la precisión de las estimaciones del efecto y el riesgo de sesgo de publicación, como se describe en la Sección 12.2.2. El sistema GRADE vincula una evaluación de la calidad del conjunto de las pruebas para cada resultado individual.

El enfoque GRADE especifica cuatro niveles de calidad (Tabla 12.2a). La calificación de calidad más alta es para las pruebas provenientes de los ensayos clínicos aleatorios. Sin embargo, los revisores pueden disminuir esta calificación de las pruebas de calidad de los ensayos clínicos aleatorios a moderada, baja e incluso muy baja, según la presencia de los cinco factores de la Tabla 12.2.b. Habitualmente la calificación de la calidad disminuirá un nivel por cada factor, hasta un máximo de tres niveles para todos los factores. Si existen problemas muy graves para cualquier factor (p.ej. cuando se evaluaron las limitaciones en el diseño y la implementación, en ningún estudio hubo ocultación ni cegamiento y más del 50% de los pacientes se perdieron durante el seguimiento) las pruebas de los ensayos clínicos aleatorios pueden disminuir dos niveles debido a este factor solo.

Generalmente los revisores calificarán las pruebas de estudios observacionales sólidos como de baja calidad. Sin embargo, si dichos estudios producen grandes efectos y no hay un sesgo evidente que explique estos efectos, los revisores pueden calificar las pruebas como moderadas o incluso (si el efecto es suficientemente grande) de alta calidad (Tabla 12.2.c). El nivel de muy baja calidad incluye, pero no se limita a, estudios con problemas críticos y observaciones clínicas no sistemáticas (p.ej. series de casos o informes de casos).

Tabla 12.2.a: Niveles de calidad del conjunto de las pruebas en el enfoque GRADE

Metodología subyacente	Calificación de calidad
Ensayos clínicos aleatorios; o estudios observacionales con una calificación dos veces más alta.	Alta
Ensayos clínicos aleatorios con una calificación más baja; o estudios observacionales con una calificación más alta.	Moderada
Ensayos clínicos aleatorios con una calificación dos veces más baja; o estudios observacionales.	Baja
Ensayos clínicos aleatorios con una calificación tres veces más baja; o estudios observacionales con una calificación más baja; o series de casos/informes de casos.	Muy baja

Tabla 12.2.b: Factores que pueden disminuir el nivel de calidad de un conjunto de pruebas

1. Limitaciones en el diseño y la implementación de los estudios disponibles que indican una alta probabilidad de sesgo.
2. Falta de direccionalidad de las pruebas (poblaciones, intervención, control y resultados indirectos).
3. Heterogeneidad inexplicada o inconsistencia de los resultados (incluidos los problemas con los análisis de subgrupos).
4. Imprecisión de los resultados (intervalos de confianza amplios).
5. Alta probabilidad de sesgo de publicación.

Tabla 12.2.c: Factores que pueden aumentar el nivel de calidad de un conjunto de pruebas

1. Gran magnitud del efecto.
2. Todos los posibles factores de confusión reducirían un efecto demostrado o indicarían un efecto espurio cuando los resultados no muestran un efecto.
3. Gradiente de dosis respuesta.

12.2.2 Factores que disminuyen el nivel de calidad de un conjunto de pruebas

A continuación se describen en más detalle los cinco motivos para disminuir la calificación de la calidad de un conjunto de pruebas para un resultado específico (Tabla 12.2.b). En cada caso si se encuentra un motivo para disminuir la calificación de las pruebas, el mismo se debe clasificar como “grave” (que disminuye la calificación de la calidad en un nivel) o “muy grave” (que disminuye la calificación en dos niveles).

1. Limitaciones en el diseño y la implementación: la confianza en una estimación del efecto disminuye si los estudios presentan limitaciones importantes que es probable que den lugar a una evaluación sesgada del efecto de la intervención. Para los ensayos clínicos aleatorios estas limitaciones metodológicas incluyen falta de ocultación de la asignación, falta de cegamiento (particularmente con resultados subjetivos altamente susceptibles a una

evaluación sesgada), grandes pérdidas durante el seguimiento, ensayos clínicos aleatorios interrumpidos de manera temprana debido a beneficio o informe selectivo de resultados. El Capítulo 8 proporciona una discusión detallada de las evaluaciones del riesgo de sesgo a nivel de estudio en el contexto de una revisión Cochrane y propone un enfoque para evaluar el riesgo de sesgo para un resultado a través de los estudios como "bajo riesgo de sesgo", "riesgo de sesgo incierto" y "alto riesgo de sesgo" (Capítulo 8, Sección 8.7). Estas evaluaciones deben formar parte de manera directa del presente factor. En particular, "bajo riesgo de sesgo" indicaría "ninguna limitación"; "riesgo de sesgo incierto" indicaría "ninguna limitación" o "limitación grave" y "alto riesgo de sesgo" indicaría "limitación grave" o "limitación muy grave". Los revisores deben utilizar su sentido común para decidir entre las categorías alternativas, según la probable magnitud de los sesgos potenciales.

Cada estudio que aborde un resultado particular diferirá, en alguna medida, en cuanto al riesgo de sesgo. Los revisores deben hacer una valoración general de si la calidad de las pruebas para un resultado justifica la disminución de la calificación sobre la base de las limitaciones del estudio. La evaluación de las limitaciones del estudio se debe aplicar a los estudios que contribuyan con resultados a la tabla "Resumen de los hallazgos", en lugar de a todos los estudios que pudieran potencialmente incluirse en el análisis. En el Capítulo 8 (Sección 8.8.3) se ha indicado que el análisis principal se debe limitar a los estudios con bajo (o bajo e incierto) riesgo de sesgo.

La Tabla 12.2.d presenta las valoraciones que se deben hacer a partir de las evaluaciones del riesgo de sesgo y hasta las valoraciones acerca de las limitaciones del estudio para cada resultado incluido en la tabla "Resumen de los hallazgos". Es posible alcanzar una calificación de alta calidad de las pruebas sólo cuando la mayoría de las pruebas proviene de estudios que cumplen los criterios de bajo riesgo de sesgo. Por ejemplo, de los 22 ensayos clínicos que abordan la repercusión de los betabloqueantes sobre la mortalidad en pacientes con insuficiencia cardíaca, la mayoría probable o ciertamente utilizó la ocultación del cegamiento, en todos al menos algún grupo clave estaba cegado y el seguimiento de los pacientes asignados al azar fue casi completo (Brophy 2001). La calidad de las pruebas pudiera disminuirse en un nivel cuando la mayoría de las pruebas proviene de estudios individuales con limitaciones fundamentales para un criterio o con algunas limitaciones para múltiples criterios. Por ejemplo, no es posible confiar en que en los pacientes con paludismo por falciparum la amodiaquina y la sulfadoxina-pirimetamina juntas reducen los fracasos del tratamiento comparadas con sulfadoxina-pirimetamina, porque la ventaja aparente de la sulfadoxina-pirimetamina fue sensible a las suposiciones con respecto a la tasa de eventos en los pacientes que se perdieron durante el seguimiento (> 20% de pérdidas durante el seguimiento en dos de tres estudios) (McIntosh 2005). Un ejemplo de limitaciones muy graves, que justifican una disminución de dos niveles en la calificación, lo proporcionan las pruebas del tratamiento quirúrgico versus conservador en los pacientes con prolapso del disco lumbar (Gibson 2007). No hay certidumbre de los beneficios de la cirugía en la reducción de los síntomas después de un año o más, ya que un ensayo clínico incluido en el análisis tuvo un ocultamiento inadecuado de la asignación y el cirujano, que no estaba cegado a la intervención, evaluó el resultado mediante una calificación bruta.

2. Falta de direccionalidad de las pruebas: hay dos tipos de falta de direccionalidad relevantes. Primero, una revisión que compare la efectividad de intervenciones alternativas (por ejemplo A y B) puede encontrar que hay ensayos clínicos aleatorios disponibles, pero que han comparado A con placebo y B con placebo. Por lo tanto, las pruebas están limitadas a las comparaciones indirectas entre A y B.

Segundo, una revisión puede encontrar ensayos clínicos aleatorios que cumplan los criterios de elegibilidad pero aborden una versión limitada de la pregunta de revisión principal en cuanto a población, intervención, comparador o resultados. Por ejemplo, suponga que en una revisión que aborda una intervención para la prevención secundaria de la cardiopatía coronaria la mayoría de los estudios identificados se realizaron en personas que también tenían diabetes. Luego, las pruebas pueden considerarse indirectas en relación con la

pregunta más amplia de interés porque la población está limitada a personas con diabetes. El escenario opuesto se puede aplicar de la misma manera: una revisión que aborde el efecto de estrategias preventivas para la cardiopatía coronaria en personas con diabetes puede considerar que los ensayos clínicos en personas sin diabetes proporcionan pruebas relevantes, aunque indirectas. Lo anterior pudiera ser particularmente probable si los investigadores han realizado pocos, o ningún, ensayos clínicos aleatorios en la población objetivo (p.ej. personas con diabetes). Otras fuentes de falta de direccionalidad pueden surgir de las intervenciones estudiadas (p.ej. si en todos los estudios incluidos expertos altamente especializados implementaron una intervención técnica en centros especializados, las pruebas sobre los efectos de la intervención fuera de estos centros pueden ser indirectas), los comparadores utilizados (p.ej. si los grupos control recibieron una intervención que es menos efectiva que el tratamiento estándar en la mayoría de los ámbitos) y los resultados evaluados (p.ej. falta de direccionalidad debido a resultados sustitutos cuando los datos sobre resultados importantes de los pacientes no están disponibles, o cuando los investigadores solicitaron datos sobre la calidad de vida pero sólo se informaron los síntomas). Los revisores deben hacer valoraciones transparentes cuando consideren que se justifica la disminución de la calificación debido a las diferencias en los efectos anticipados en el grupo de interés principal.

3. Heterogeneidad inexplicada o inconsistencia de los resultados: cuando los estudios producen diferentes estimaciones del efecto (heterogeneidad o variabilidad en los resultados) los investigadores deben buscar explicaciones sólidas para dicha heterogeneidad. Por ejemplo, los fármacos pueden tener efectos relativamente mayores en las poblaciones más enfermas o cuando se dan en dosis más altas. Una discusión detallada de la heterogeneidad y su investigación se proporciona en el Capítulo 9 (Secciones 9.5 y 9.6). Si existe un modificador importante, con pruebas sólidas de que los resultados importantes son diferentes en diferentes subgrupos (idealmente preespecificados), es posible considerar una tabla “Resumen de los hallazgos” por separado para una población separada. Por ejemplo, sería posible utilizar una tabla “Resumen de los hallazgos” separada para la endarterectomía carotídea en pacientes sintomáticos con un alto grado de estenosis en los cuales la intervención es, en manos de cirujanos expertos, beneficiosa (Cina 2000), y otra (si se considera que vale la pena) para pacientes asintomáticos con un grado moderado de estenosis en los cuales la cirugía no es beneficiosa (Chambers 2005). Cuando existe heterogeneidad y la misma afecta la interpretación de los resultados, pero los revisores no pueden identificar una explicación plausible, disminuye la calidad de las pruebas.
4. Imprecisión de los resultados: cuando los estudios incluyen pocos participantes y pocos eventos y por lo tanto los intervalos de confianza son amplios, los revisores pueden disminuir su calificación de la calidad de las pruebas. Los intervalos de confianza incluidos en la tabla “Resumen de los hallazgos” proporcionarán a los lectores información que les permita hacer, en cierta medida, su propia calificación de la precisión.
5. Alta probabilidad de sesgo de publicación: la calificación de la calidad del nivel de las pruebas puede disminuir si los investigadores no pueden informar estudios (generalmente los que no muestran un efecto: sesgo de publicación) ni resultados (generalmente los que pueden ser perjudiciales o para los cuales no se observó un efecto: sesgo de informe selectivo de resultados) sobre la base de los hallazgos. El informe selectivo de los resultados se evalúa a nivel de estudio como parte de la evaluación del riesgo de sesgo (ver Capítulo 8, Sección 8.13), de manera que para los estudios que contribuyen al resultado en la tabla “Resumen de los hallazgos” este aspecto se aborda en el factor 1 descrito anteriormente (limitaciones en el diseño y la implementación). Si un gran número de estudios incluidos en la revisión no contribuyen a un resultado, o si hay pruebas de sesgo de publicación, la calificación de la calidad de las pruebas puede disminuir. El Capítulo 10 proporciona una discusión detallada de los sesgos de informe, incluido el sesgo de publicación y cómo se puede evitar en una revisión Cochrane. Una situación prototipo que puede provocar sospechas de sesgo de publicación es cuando las pruebas publicadas incluyen varios ensayos clínicos pequeños que están financiados por la industria (Bhandari 2004). Por ejemplo, 14 ensayos clínicos de flavonoides en pacientes con hemorroides han mostrado grandes beneficios aparentes, pero en total reclutaron sólo 1432 pacientes (es decir, cada ensayo clínico reclutó relativamente pocos pacientes) (Alonso-Coello 2006). La importante participación de los patrocinadores en la

mayoría de estos ensayos clínicos hace surgir inquietudes acerca de si existen ensayos clínicos no publicados que indiquen que no hay beneficios.

Un conjunto de pruebas particular puede tener problemas asociados con más de uno de los cinco factores mencionados anteriormente y, mientras mayores son los problemas, menor es la calificación que debe obtener la calidad de las pruebas. Es posible imaginar una situación en la cual hubiera ensayos clínicos aleatorios disponibles pero todas o casi todas estas limitaciones estuvieran presentes y de una manera importante. Lo anterior pudiera dar lugar a una calificación muy baja de la calidad de las pruebas.

Tabla 12.2.d: Guía adicional para el factor 1 (de 5) en una evaluación GRADE: de las evaluaciones del riesgo de sesgo a las valoraciones acerca de las limitaciones del estudio para los resultados principales

Riesgo de sesgo	Entre los estudios	Interpretación	Consideraciones	Evaluación GRADE de las limitaciones del estudio
Bajo riesgo de sesgo.	La mayor parte de la información proviene de estudios con bajo riesgo de sesgo.	Poco probable que el posible sesgo modifique de manera importante los resultados.	Ninguna limitación evidente.	Ninguna limitación importante, no disminuir la calificación.
Riesgo incierto de sesgo.	La mayor parte de la información proviene de estudios con riesgo de sesgo bajo o incierto.	Posible sesgo que hace surgir algunas dudas acerca de los resultados.	Es poco probable que las posibles limitaciones disminuyan la confianza en la estimación del efecto.	Ninguna limitación importante, no disminuir la calificación.
			Es probable que las posibles limitaciones disminuyan la confianza en la estimación del efecto.	Limitaciones importantes, disminuir la calificación un nivel.
Alto riesgo de sesgo.	La proporción de información proveniente de estudios con alto riesgo de sesgo es suficiente para afectar la interpretación de los resultados.	Posible sesgo que debilita de manera importante la confianza en los resultados.	Limitación fundamental en un criterio, o algunas limitaciones en múltiples criterios, suficientes para disminuir la confianza en la estimación del efecto.	Limitaciones importantes, disminuir la calificación un nivel.
			Limitaciones fundamentales en uno o más criterios, suficientes para disminuir de manera importante la confianza en la estimación del efecto.	Limitaciones muy importantes, disminuir la calificación dos niveles.

12.2.3 Factores que aumentan el nivel de calidad de un conjunto de pruebas

Aunque los estudios observacionales y los ensayos clínicos aleatorios a los que se les disminuyó la calificación generalmente producirán una calificación baja de la calidad de las pruebas, hay circunstancias poco frecuentes en las cuales los revisores pudieran "aumentar" la calificación de dichas pruebas a moderada o incluso alta calidad (Tabla 12.2.c).

1. En pocas ocasiones, cuando los estudios observacionales metodológicamente bien realizados producen estimaciones grandes, consistentes y precisas de la magnitud del efecto de una intervención, es posible confiar particularmente en los resultados. Lo anterior puede ser el caso cuando existe un efecto grande (p.ej. CR > 2 o CR < 0,5) en ausencia de posibles factores de confusión, o un efecto muy grande (p.ej. CR > 5 o CR < 0,2) en estudios sin amenazas importantes para la validez. En estas situaciones, aunque es probable que los estudios observacionales hayan proporcionado una sobrestimación del efecto verdadero, las debilidades del diseño del estudio no pueden explicar todos los beneficios aparentes observados. Por lo tanto, a pesar de las reservas con respecto al diseño observacional del estudio, los revisores confían en que el efecto existe. La magnitud del efecto en estos estudios puede hacer que la calificación asignada a la calidad de las pruebas cambie de baja a moderada (si el efecto es grande en ausencia de otras limitaciones metodológicas). Por ejemplo, un metanálisis de estudios observacionales mostró que los cascos de bicicleta reducen en gran medida el riesgo de lesiones en la cabeza para los ciclistas (odds ratio [OR] 0,31; IC del 95%: 0,26 a 0,37) (Thompson 2000). Este efecto grande, ante la falta de un sesgo evidente que pudiera crear dicha asociación, indica una calificación moderada de la calidad de las pruebas.

2. En ocasiones todos los sesgos posibles de los estudios observacionales o aleatorios pueden actuar juntos para sobrestimar un efecto evidente de una intervención. Por ejemplo, si sólo los pacientes más enfermos reciben una intervención o exposición experimental, pero todavía responden mejor, es probable que el efecto de la intervención o la exposición real sea mayor que lo que indican los datos. Como muestra, una revisión sistemática rigurosa de estudios observacionales que incluyeron en total 38 millones de pacientes demostró tasas de muerte mayores en los hospitales privados con fines de lucro comparados con los hospitales privados sin fines de lucro (Devereaux 2004). Un sesgo posible se relaciona con la diferente gravedad de la enfermedad en los pacientes de los dos tipos de hospitales. Sin embargo, es probable que los pacientes de los hospitales sin fines de lucro estuvieran más enfermos que los de los hospitales con fines de lucro. Por lo tanto, hasta donde se conoce la presencia de factores de confusión residuales, los mismos sesgarían los resultados en contra de los hospitales sin fines de lucro. El segundo sesgo probable fue la posibilidad de que un mayor número de pacientes con una cobertura excelente de seguro privado puede dar lugar a que el hospital tenga más recursos y un efecto de "desbordamiento" que beneficiaría a los que no tiene dicha cobertura. Como es probable que los hospitales con fines de lucro ingresen una mayor proporción de estos pacientes bien asegurados, nuevamente aparece el sesgo en contra de los hospitales sin fines de lucro. Como todos los sesgos posibles disminuirían el efecto demostrado de la intervención, es posible considerar que la calidad de las pruebas de estos estudios observacionales es moderada, en lugar de baja. Ocurre una situación similar cuando los estudios observacionales no pueden demostrar una asociación pero todos los sesgos posibles aumentarían el efecto de una intervención. Esta situación surge generalmente cuando se exploran efectos perjudiciales evidentes. Por ejemplo, debido a que el fármaco hipoglucémico fenformina produce acidosis láctica, se sospecha que el agente relacionado metformina presenta la misma toxicidad. No obstante, estudios observacionales muy grandes no han podido demostrar una asociación (Salpeter 2007). Como es probable que los médicos estén más alerta hacia la acidosis láctica en presencia del agente y sobreinformen su ocurrencia, es posible considerar que estas pruebas de calidad moderada o incluso alta niegan una relación causal entre las dosis terapéuticas típicas de metformina y la acidosis láctica.

3. La presencia de un gradiente de dosis respuesta también puede aumentar la confianza en los hallazgos de los estudios observacionales y mejorar así la calificación asignada a la calidad de las pruebas. Por ejemplo, la confianza en el resultado de estudios observacionales que muestran un aumento del riesgo de hemorragia en pacientes que tienen niveles de coagulación supraterapéuticos aumenta por la observación de que hay un gradiente de dosis

respuesta entre niveles mayores de la razón internacional normalizada (RIN) y el aumento del riesgo de hemorragia (Levine 2004).

12.3 Aspectos de la aplicabilidad

12.3.1 La función del revisor

"Siempre se necesita un acto de fe cuando se aplican los hallazgos de algún estudio a la población general" o a una persona específica. "Cuando se realiza este salto, siempre se debe establecer un equilibrio entre hacer generalizaciones amplias justificables y ser demasiado conservador en la conclusiones" (Friedman 1985).

Para abordar adecuadamente el grado en el cual una revisión es relevante para los objetivos para los que se realiza ("direccionalidad") hay ciertas cosas que el revisor debe hacer, y ciertas cosas que el usuario de la revisión debe hacer. Aquí se discute qué puede hacer el revisor para ayudar al usuario. Los revisores Cochrane deben ser muy claros sobre la población, la intervención y los resultados que intentan abordar. El Capítulo 11 (Sección 11.5.2) hace énfasis en un paso fundamental que tradicionalmente no ha sido parte de las revisiones Cochrane: la especificación de todos los resultados importantes para los pacientes relevantes para las estrategias de investigación bajo comparación.

Con respecto a los factores del participante y la intervención los revisores necesitan hacer hipótesis *a priori* acerca de los posibles modificadores del efecto, y luego examinar dichas hipótesis. Si encuentran efectos evidentes de subgrupos, deben decidir finalmente si estos efectos son creíbles o no (Oxman 2002). Las diferencias entre subgrupos, particularmente aquellas que corresponden a diferencias entre los estudios, se deben interpretar con precaución. Es inevitable alguna variación entre los subgrupos debido al azar, por lo que a menos que existan pruebas sólidas de una interacción los revisores no deben suponer que exista un efecto de subgrupos. Si a pesar de todas las precauciones los revisores consideran que los efectos de subgrupos son fidedignos, deben realizar metanálisis por separado para los subgrupos relevantes, y producir tablas "Resumen de los hallazgos" separadas para estos subgrupos.

El usuario de la revisión tendrá el reto de "individualizar" los hallazgos. Por ejemplo, incluso si los efectos relativos son similares entre los subgrupos, los efectos absolutos diferirán según el riesgo inicial. Los revisores pueden ayudar a proporcionar esta información al identificar grupos de personas identificables con riesgos variables en las tablas "Resumen de los hallazgos", como se discute en el Capítulo 11 (Sección 11.5.5). Los usuarios pueden identificar entonces a los pacientes que están ante ellos como pertenecientes a un grupo de riesgo particular, y evaluar por lo tanto la probable magnitud del beneficio o el daño.

Otra decisión que deben tomar los usuarios es si los pacientes que están ante ellos son tan diferentes de los incluidos en los estudios que no es posible utilizar los resultados de la revisión sistemática y el metanálisis. Los revisores pueden señalar que, en lugar de aplicar rígidamente los criterios de inclusión y exclusión de los estudios, es mejor preguntarse si hay razones importante por las cuales no se deben aplicar las pruebas a un paciente particular (Guyatt 1994). En algunas ocasiones los revisores pueden ayudar a quienes toman decisiones médicas al identificar variaciones importantes que pudieran limitar la aplicabilidad de los resultados (Schünemann 2006a), incluidas las variaciones biológicas y culturales y las variaciones en el cumplimiento con una intervención.

Al abordar estos aspectos, es posible que los revisores no estén al tanto ni aborden la gran cantidad de diferencias en las circunstancias en todo el mundo. Sin embargo, pueden abordar las diferencias de importancia reconocida para muchas personas y, de manera importante, deben evitar suponer que las circunstancias de otras personas son las mismas que las propias al discutir los resultados y establecer conclusiones.

12.3.2 Variación biológica

Los aspectos de la variación biológica que los revisores deben considerar incluyen las divergencias en la fisiopatología (p.ej. las diferencias biológicas entre las mujeres y los hombres que es probable que afecten la respuesta al tratamiento) y las divergencias en el agente causal (p.ej. para las enfermedades infecciosas como el paludismo).

12.3.3 Variación en el contexto y la cultura

Algunas intervenciones, particularmente las no farmacológicas, pueden funcionar en algunos contextos pero no en otros; la situación se ha descrito como un programa por interacción de contexto (Hawe 2004). Los factores de contexto pudieran estar relacionados con la organización huésped en la cual se ofrece la intervención, como la pericia, la experiencia y los patrones morales del personal que se espera que realice la intervención, las prioridades contradictorias para la atención del personal, los recursos locales como los servicios y facilidades disponibles para el programa y el estado o importancia dados al programa por la organización huésped. Los aspectos de contexto más amplios pudieran incluir características del sistema dentro del cual opera la organización huésped, como la estructura de financiamiento o pago para los prestadores de atención sanitaria. Los factores de contexto también se pueden relacionar con las características de los servicios para el grupo o la población objetivo (dichos aspectos incluyen la diversidad cultural y de idiomas, la posición socioeconómica, el ámbito urbano/rural), lo que puede significar que se desarrolle un estilo particular de atención o relación entre los proveedores del servicio y los consumidores que coincida o no con los valores y la tecnología del programa. Durante muchos años se han reconocido estos aspectos (pero no se han especificado claramente) cuando quienes toman las decisiones han argumentado que los resultados de las revisiones de las pruebas de otros países no se aplican a su propio país.

Aunque algunos programas/intervenciones se han transferido de un contexto a otro y se han observado beneficios, en otros casos no ha sido así (Resnicow 1993, Lumley 2004). Los revisores deben tener precaución cuando hacen generalizaciones de un contexto a otro. También deberían señalar la presencia (o no) de información relacionada con el contexto en los estudios de intervención, cuando esta información esté disponible (Hawe 2004).

12.3.4 Variación en el cumplimiento

La variación en el cumplimiento de quienes reciben y quienes proporcionan la atención puede limitar la aplicabilidad de los resultados. Las diferencias predecibles en el cumplimiento se pueden deber a divergencias en las condiciones económicas o las actitudes, que hacen que algunas formas de atención no sean accesible o factibles en algunos ámbitos como los países en desarrollo (Dans 2007). No se debería suponer que altos niveles de cumplimiento en ensayos clínicos aleatorios cuidadosamente monitorizados se traducirán en niveles similares de cumplimiento en la práctica habitual.

12.3.5 Variación en los valores y preferencias

Las decisiones de tratamiento incluyen equilibrar los beneficios y las desventajas de las estrategias de tratamiento propuestas. La selección correcta puede diferir en personas con diferentes valores y preferencias, y depende del médico asegurar que las decisiones seas consistentes con los valores y preferencias de los pacientes. En la Sección 12.7 se describe cómo los revisores pueden ayudar a este proceso.

12.4 Interpretación de los resultados de los análisis estadísticos

12.4.1 Intervalos de confianza

Los resultados de los estudios individuales y los metanálisis se informan con una estimación puntual junto con un intervalo de confianza asociado. Por ejemplo, “El odds ratio fue 0,75 con un intervalo de confianza del 95% de 0,70 a 0,80”. La estimación puntual (0,75) es la mejor

conjetura acerca de la magnitud y dirección del efecto de la intervención experimental comparada con la intervención control. El intervalo de confianza describe la incertidumbre inherente a esta estimación y describe un rango de valores dentro del cual se puede estar razonablemente seguro que realmente se encuentra el verdadero efecto. Si el intervalo de confianza es relativamente estrecho (p.ej. 0,70 a 0,80), el tamaño del efecto se conoce con precisión. Si el intervalo es más amplio (p.ej. 0,60 a 0,93) la incertidumbre es mayor, aunque aún hay suficiente precisión para tomar decisiones acerca de la utilidad de la intervención. Los intervalos muy amplios (p.ej. 0,50 a 1,10) indican que se conoce poco acerca del efecto y que se necesita información adicional.

A menudo un intervalo de confianza del 95% se interpreta como que indica un rango dentro del cual se puede estar 95% seguro de que se encuentra el verdadero efecto. Este planteamiento es una interpretación poco exacta, pero es útil como guía general. La interpretación estrictamente correcta de un intervalo de confianza se basa en la noción hipotética de considerar los resultados que se obtendrían si el estudio se repitiera muchas veces. Si un estudio se repitiera de forma indefinida y en cada ocasión se calculara un intervalo de confianza del 95%, entonces en el 95% de estos intervalos estaría contenido el verdadero efecto.

La amplitud del intervalo de confianza para un estudio individual depende en gran medida del tamaño de la muestra. Los estudios más grandes tienden a producir estimaciones más precisas de los efectos (y por lo tanto, a tener intervalos de confianza más estrechos) que los estudios más pequeños. Para los resultados continuos, la precisión depende también de la variabilidad en las mediciones del resultado (la desviación estándar de las mediciones entre los individuos); para los resultados dicotómicos lo anterior depende del riesgo del riesgo del evento, y para los resultados de tiempo hasta el evento depende del número de eventos observados. Todas estas cantidades se utilizan para el cálculo de los errores estándar de las estimaciones del efecto, de los cuales se deriva el intervalo de confianza.

La amplitud de un intervalo de confianza para un metanálisis depende de la precisión de las estimaciones de los estudios individuales y del número de estudios combinados. Además, para los modelos de efectos aleatorios la precisión disminuirá cuando aumenta la heterogeneidad y los intervalos de confianza se ampliarán en consecuencia (ver Capítulo 9, Sección 9.5.4). Cuando se agregan más estudios al metanálisis habitualmente la amplitud del intervalo de confianza disminuye. Sin embargo, si los estudios adicionales aumentan la heterogeneidad en el metanálisis y se utiliza un modelo de efectos aleatorios, es posible que la amplitud del intervalo de confianza aumente.

Los intervalos de confianza y las estimaciones puntuales tienen diferentes interpretaciones en los modelos de efectos fijos y aleatorios. Mientras que la estimación del efecto fijo y su intervalo de confianza abordan la pregunta “¿cuál es la mejor estimación (única) del efecto?”, las estimaciones de los efectos aleatorios suponen que hay una distribución de los efectos, y la estimación y su intervalo de confianza abordan la pregunta “¿cuál es la mejor estimación del efecto promedio?”.

Un intervalo de confianza se puede informar para cualquier nivel de confianza (aunque se informan con mayor frecuencia para el 95% y algunas veces para el 90% o el 99%). Por ejemplo, el odds ratio de 0,80 se pudiera informar con un intervalo de confianza del 80% de 0,73 a 0,88; un intervalo de confianza del 90% de 0,72 a 0,89; y un intervalo de confianza del 95% de 0,70 a 0,92. Cuando el nivel de confianza aumenta el intervalo de confianza se amplía.

Existe una correspondencia lógica entre el intervalo de confianza y el valor de p (ver Sección 12.4.2). El intervalo de confianza del 95% para un efecto excluirá el valor nulo (como un odds ratio de 1,0 o una diferencia de riesgo de 0) si y sólo si la prueba de significación produce un valor de p menor de 0,05. Si el valor de p es exactamente 0,05 el límite superior o inferior del intervalo de confianza del 95% estará en el valor nulo. De manera similar, el intervalo de

confianza del 99% excluirá el valor nulo si y sólo si la prueba de significación produce un valor de p menor de 0,01.

Juntos, la estimación puntual y el intervalo de confianza proporcionan información para evaluar la utilidad clínica de la intervención. Por ejemplo, suponga que se evalúa un tratamiento que reduce el riesgo de un evento y se decide que el mismo sería útil sólo si reduce el riesgo de un evento del 30% en al menos 5 puntos porcentuales, a 25% (estos valores dependerán del escenario clínico y el resultado específicos). Si el metanálisis produce una estimación del efecto de una reducción de 10 puntos porcentuales con un intervalo de confianza del 95% estrecho, por ejemplo, del 7% al 13%, es posible concluir que el tratamiento fue útil porque la estimación puntual y el rango completo del intervalo exceden el criterio de una reducción del 5% para la utilidad clínica. Sin embargo, si el metanálisis informó la misma reducción del riesgo del 10% pero con un intervalo de confianza más amplio, por ejemplo, del 2% al 18%, aunque aún se concluiría que la mejor estimación del efecto del tratamiento es que es útil, no es posible tener una confianza similar, ya que no se ha excluido la posibilidad de que el efecto pueda estar entre el 2% y el 5%. Si el intervalo de confianza fuera aún más amplio e incluyera el valor nulo de una diferencia del 0%, no se habrá excluido la posibilidad de que el tratamiento tenga un efecto en cualquier dirección, y sería necesario ser más escéptico en las conclusiones.

Los intervalos de confianza con diferentes niveles de confianza pueden demostrar que existen pruebas diferenciales para diferentes grados de beneficio o daño. Por ejemplo, es posible informar los mismos resultados del análisis (i) con una confianza del 95% de que la intervención no causa daños; (ii) con una confianza del 90% de que tiene algún efecto; y (iii) con una confianza del 80% de que tiene importantes beneficios para el paciente. Estos elementos pueden indicar utilidad de la intervención y necesidad de investigaciones adicionales.

Los revisores pueden utilizar el mismo enfoque general para concluir que una intervención *no* es útil. A partir del ejemplo anterior donde el criterio para una diferencia mínima importante para el paciente es una diferencia de riesgo del 5%, una estimación del efecto del 2% con un intervalo de confianza del 1% al 4% indica que la intervención no es útil.

12.4.2 Valores de p y significación estadística

Un valor de p es la probabilidad de obtener el efecto observado (o uno mayor) bajo la “hipótesis nula”, la cual en el contexto de la revisión Cochrane es una suposición de “ningún efecto de la intervención” o “ninguna diferencia en el efecto de la intervención entre los estudios” (sin heterogeneidad). Por lo tanto, un valor de p muy pequeño indica que es muy poco probable que el efecto observado haya aparecido solamente por azar, por lo que aporta pruebas contra la hipótesis nula. Ha sido una práctica frecuente interpretar el valor de p al examinar si es más pequeño que un valor umbral particular. En particular los valores de p menores de 0,05 se informan a menudo como “estadísticamente significativos”, y se interpretan como suficientemente pequeños como para rechazar la hipótesis nula. Sin embargo, el umbral de 0,05 es un valor arbitrario que se ha utilizado frecuentemente en las investigaciones médicas y psicológicas, en gran parte porque los valores de p se determinaron al comparar las pruebas estadísticas contra las tabulaciones de puntos porcentuales específicos de las distribuciones estadísticas. RevMan, como otros programas estadísticos, informa los valores exactos de p. Si los revisores deciden presentar un valor de p con los resultados de un metanálisis, deben informar un valor exacto de p junto con el intervalo de confianza del 95%.

En RevMan se proporcionan dos valores de p. Uno se relaciona con el efecto resumen en el metanálisis y proviene de una prueba Z de la hipótesis nula de ningún efecto (o ningún efecto promedio en un metanálisis de efectos aleatorios). El otro se relaciona con la heterogeneidad entre los estudios y proviene de una prueba de ji cuadrado de la hipótesis nula de que no existe heterogeneidad (ver Capítulo 9, Sección 9.5.2).

Para las pruebas de un efecto resumen, el cálculo del valor de p incluye la estimación del efecto y el tamaño de la muestra (o, más estrictamente, la precisión de la estimación del efecto). Cuando el tamaño de la muestra aumenta, se reduce el rango de efectos posibles que pudieran ocurrir por azar. De manera correspondiente, la significación estadística de un efecto de una magnitud particular será mayor (el valor de p será más pequeño) en un estudio más grande que en un estudio más pequeño.

Con frecuencia los valores de p se interpretan erróneamente de dos maneras. Primero, un valor de p moderado o grande (p.ej. mayor de 0,05) se puede interpretar erróneamente como prueba de que "la intervención no tiene efecto". Existe una diferencia importante entre este planteamiento y la interpretación correcta de que "no hay pruebas sólidas de que la intervención tiene un efecto". Para evitar esta interpretación errónea los revisores deben examinar siempre la estimación del efecto y su intervalo de confianza del 95% junto con el valor de p. En estudios o metanálisis pequeños es frecuente incluir ningún efecto de la intervención y un efecto significativo, en el rango de efectos contenidos en el intervalo de confianza. Se les aconseja a los revisores no describir resultados como "no estadísticamente significativo" o "no significativo".

La segunda interpretación errónea es suponer que un resultado con un valor de p pequeño para la estimación del efecto resumen implica que una intervención tiene un beneficio importante. Esta interpretación errónea es más probable que ocurra en estudios grandes como los metanálisis donde se agrupan datos de docenas de estudios y miles de participantes. El valor de p aborda la interrogante de si el efecto de la intervención es exactamente 0; no examina si el efecto es de una magnitud importante para los posibles receptores de la intervención. En un estudio grande un valor de p pequeño puede representar la detección de un efecto trivial. Nuevamente, la inspección de la estimación puntual y el intervalo de confianza ayudan a las interpretaciones correctas (ver Sección 12.4.1).

12.5 Interpretación de los resultados dicotómicos (incluidos los números necesarios a tratar)

12.5.1 Reducciones relativas y absolutas del riesgo

Los médicos pueden estar más inclinados a prescribir una intervención que reduce el riesgo de muerte en el 25% que una que reduce el riesgo de muerte en un punto porcentual, aunque ambas presentaciones de las pruebas pueden relacionarse con el mismo beneficio (es decir, una reducción en el riesgo del 4% al 3%). La primera se refiere a la reducción *relativa* en el riesgo y la segunda a la reducción *absoluta* en el riesgo. Como se describe en el Capítulo 9 (Sección 9.2.2), existen numerosas medidas para comparar los resultados dicotómicos en dos grupos. Habitualmente para realizar los metanálisis se utilizan los cocientes de riesgos (CR), los odds ratios (OR) o las diferencias de riesgos (DR), pero hay muchas formas alternativas de expresar los resultados.

La reducción del riesgo relativo (RRR) es una forma conveniente de reexpresar un cociente de riesgos como una reducción porcentual:
RRR = 100% × (1 – RR).
Por ejemplo, una reducción de 0,75 se traduce en una reducción del riesgo relativo del 25% en el ejemplo anterior.

A menudo, a la diferencia de riesgo se le llama reducción del riesgo absoluto (RRA), y se puede presentar como un porcentaje (por ejemplo, 1%), como un decimal (por ejemplo, 0,01) o como recuentos (por ejemplo, 10 de 1000). Una transformación sencilla de la diferencia de riesgo conocida como número necesario a tratar (NNT) es una alternativa frecuente para presentar la misma información. En la Sección 12.5.2 se discuten los NNT y en la Sección 12.5.3 se consideran diferentes opciones para presentar los efectos absolutos.

Luego se describen los cálculos para obtener estos números a partir de los resultados de los estudios individuales y de los metanálisis.

12.5.2 Más acerca del número necesario a tratar (NNT)

El número necesario a tratar (NNT) se define como el número esperado de personas que necesitarían recibir la intervención experimental en lugar de la intervención control para que en una persona adicional ocurra o se evite un evento en un período de tiempo determinado. Por lo tanto, como ejemplo, un NNT de 10 se puede interpretar como "se espera que una persona (o menos) adicional presentará un evento por cada 10 participantes que reciban la intervención experimental en lugar de la intervención control en un período de tiempo determinado". Es importante tener claro que:

1. como el NNT se deriva de la diferencia de riesgo, todavía es una medida *comparativa* del efecto (experimental versus algún control) y que no es una propiedad general de una intervención única; y
2. el NNT proporciona un "valor esperado". Por ejemplo, un NNT = 10 no implica que un evento adicional *ocurrirá* en todos y cada uno de los grupos de diez personas.

El NNT se puede calcular para los efectos beneficiosos y perjudiciales y para intervenciones que causen mejoría y empeoramiento en los resultados. En todos los casos los NNT se expresan como números enteros positivos y todos los decimales se redondean. Algunos revisores utilizan el término "número necesario para dañar" (NND) cuando una intervención provoca un empeoramiento en lugar de mejoría del resultado. Sin embargo, esta frase es desagradable, errónea e inexacta (más aún, es posible interpretar fácilmente que implica el número de personas que presentará un resultado perjudicial si se les administra la intervención) y se recomienda firmemente que se eviten los términos "número necesario para dañar" y "NND". La alternativa preferida es utilizar frases como "número necesario a tratar para un resultado beneficioso adicional" (NNTB) y "número necesario a tratar para un resultado perjudicial adicional" (NNTD) para indicar la dirección del efecto.

Como el NNT se refiere a eventos, su interpretación se debe expresar cuidadosamente cuando el resultado binario es una dicotomización de un resultado basado en una escala. Por ejemplo, si el resultado es dolor medido en una escala como "ninguno, leve, moderado o intenso", el mismo se puede dicotomizar como "ninguno o leve" versus "moderado o intenso". No sería apropiado referirse al NNT a partir de estos datos como un "NNT para dolor". El mismo es un "NNT para dolor moderado o intenso".

12.5.3 Expresión de las reducciones del riesgo absoluto

Es de esperar que los usuarios de las revisiones estén influenciados por la selección de las presentaciones estadísticas de las pruebas. Hoffrage y cols. indican que las inferencias de los médicos acerca de los resultados estadísticos con más apropiadas cuando tratan con "frecuencias naturales" (número total de personas tratadas y no tratadas) (p.ej. resultados del tratamiento con una disminución de 20 de 1000 a 10 de 1000 mujeres con cáncer de mama), que cuando los efectos se presentan como porcentajes (p.ej. reducción absoluta del 1% en el riesgo de cáncer de mama) (Hoffrage 2000). Las probabilidades pueden ser más difíciles de comprender que las frecuencias, particularmente cuando los eventos son poco frecuentes. Aunque la estandarización puede ser importante para mejorar la presentación de las pruebas de investigación (y la participación en las decisiones de la atención sanitaria), las pruebas actuales indican que los consumidores de la información sanitaria comprenden mejor la presentación de las frecuencias naturales para expresar las diferencias en el riesgo absoluto. Estas pruebas proporcionan la justificación para presentar los riesgos absolutos en las tablas "Resumen de los hallazgos" como números de personas con eventos por 1000 personas que reciben la intervención.

Los cocientes de riesgos y las reducciones del riesgo relativo todavía son fundamentales porque los efectos relativos tienden a ser mucho más estables entre los grupos de riesgo comparados con los beneficios absolutos. Los revisores pueden utilizar sus propios datos para estudiar estas consistencias (Cates 1999, Smeeth 1999). Es menos probable que las

diferencias de riesgo sean consistentes a través de las tasas iniciales de eventos; por lo tanto, pocas veces son apropiadas para calcular el número necesario a tratar en las revisiones sistemáticas. Si se selecciona una medida de efecto relativa (OR o CR) para el metanálisis es necesario especificar un grupo de riesgo control como parte del cálculo de una RRA o un NNT. Es fundamental expresar el beneficio absoluto para cada grupo de riesgo clínicamente identificable y aclarar el período de tiempo para el cual se aplica. Los estudios en pacientes con diferentes grados de gravedad de la enfermedad o los estudios con diferentes duraciones del seguimiento, tendrán casi seguramente diferentes riesgos del grupo control. En estos casos los diferentes riesgos del grupo control dan lugar a diferentes RRA y NNT (excepto cuando la intervención no tiene efecto). Un enfoque recomendado es reexpresar un odds ratio o un cociente de riesgo como una variedad de NNT entre varios riesgos control supuestos (RCS) (McQuay 1997, Smeeth 1999, Sackett 2000). Los revisores deben tener en mente estas consideraciones no sólo cuando construyen su tabla "Resumen de los hallazgos", sino también en el texto de la revisión.

Por ejemplo, una revisión de anticoagulantes orales para prevenir el accidente cerebrovascular presentó la información a los usuarios mediante la descripción de los beneficios absolutos para varios riesgos iniciales (Aguilar 2005). Los revisores presentaron sus hallazgos principales como "Se debe considerar el riesgo inherente de accidente cerebrovascular cuando se decida utilizar anticoagulantes orales en pacientes con fibrilación auricular y seleccionar los que parezcan beneficiarse más de este tratamiento" (Aguilar 2005). Entre los pacientes con alto riesgo de fibrilación auricular con accidente cerebrovascular previo o ataque isquémico transitorio que presentan tasas de accidente cerebrovascular de alrededor del 12% (120 por 1000) por año, la warfarina previene cerca de 70 accidentes cerebrovasculares anualmente por cada 1000 pacientes, mientras que para los pacientes con bajo riesgo de fibrilación auricular (con una tasa de accidente cerebrovascular de cerca del 2% por año o 20 por 1000), la warfarina sólo previene 12 accidentes cerebrovasculares. Esta presentación ayuda a los usuarios a comprender la repercusión importante que tienen los riesgos típicos iniciales sobre el beneficio absoluto que se puede esperar.

12.5.4 Cálculos

El cálculo directo de una reducción del riesgo absoluto (RRA) o un número necesario a tratar (NNT) depende del estadístico resumen (odds ratio, cociente de riesgos o diferencia de riesgo) disponible en los estudios o el metanálisis. Cuando se expresan los resultados de los metanálisis, los revisores deben utilizar en los cálculos cualquier estadístico que determinen que es el resumen más apropiado para el agrupamiento (ver Capítulo 9, Sección 9.4.4.4). Aquí se presentan los cálculos para obtener la RRA como una reducción en el número de participantes por 1000. Por ejemplo, una diferencia de riesgo de -0,133 equivale a 133 participantes *menos* con el evento por 1000.

Las RRA y los NNT no se deben calcular a partir de los números totales agregados de participantes y eventos entre los ensayos clínicos. Este enfoque ignora la asignación al azar dentro de los estudios y puede producir resultados extremadamente erróneos si existe desequilibrio en la asignación al azar en alguno de los estudios.

Cuando se calculan los NNT los valores obtenidos, por convención, se redondean siempre al número entero exacto siguiente.

12.5.4.1 Cálculo del NNT a partir de la diferencia de riesgo (DR)

Los NNT se pueden calcular para estudios únicos de la manera siguiente. Es de señalar que este enfoque, aunque aplicable, sólo se debería utilizar en muy pocas ocasiones para los resultados de un metanálisis de diferencias de riesgos, ya que al realizar los metanálisis habitualmente se deben utilizar medidas de efecto relativas (CR u OR).

Un NNT se puede calcular a partir de una diferencia de riesgos como

$$\text{NNT} = \frac{1}{\text{absolute value of risk difference}} = \frac{1}{|\text{RD}|},$$

donde las barras verticales ("valor absoluto de") en el denominador indican que se debe ignorar cualquier signo de menos. Por convención el NNT se redondea al número entero siguiente. Por ejemplo, si la diferencia de riesgo es -0,12 el NNT es 9; si la diferencia de riesgo es -0,22 entonces el NNT es 5.

12.5.4.2 Cálculo de la reducción del riesgo absoluto o el NNT a partir de un cociente de riesgos (CR)

Para ayudar a la interpretación los revisores pueden calcular una reducción del riesgo absoluto o un NNT a partir de los resultados de un metanálisis de cocientes de riesgos. Para hacerlo se necesita un riesgo control supuesto (RCS). Habitualmente será apropiado hacerlo para varios RCS. El procedimiento de cálculo es el siguiente:
número menor por 1000 = 1000 x RCS x (1-CR),

$$\text{NNT} = \left|\frac{1}{\text{ACR} \times (1 - \text{RR})}\right|$$

Como ejemplo, suponga que el cociente de riesgo es CR = 0,92 y el riesgo control supuesto es RCS = 0,3 (300 por 1000). Entonces el efecto sobre el riesgo es 24 menos por 1000:

$$\text{number fewer per 1000} = 1000 \times 0.3 \times (1 - 0.92) = 24.$$

El NNT es 42:

$$\text{NNT} = \left|\frac{1}{0.3 \times (1 - 0.92)}\right| = \left|\frac{1}{0.3 \times 0.08}\right| = 41.67.$$

12.5.4.3 Cálculo de la reducción del riesgo absoluto o el NNT a partir de un odds ratio (OR)

Los revisores pueden calcular una reducción del riesgo absoluto o un NNT a partir de los resultados de un metanálisis de odds ratios. Para hacerlo se necesita un riesgo control supuesto (RCS). Habitualmente será apropiado hacerlo para varios RCS. El procedimiento de cálculo es el siguiente:

$$\text{number fewer per 1000} = 1000 \times \left(\text{ACR} - \frac{\text{OR} \times \text{ACR}}{1 - \text{ACR} + \text{OR} \times \text{ACR}}\right)$$

$$\text{NNT} = \frac{1}{\left|\text{ACR} - \frac{\text{OR} \times \text{ACR}}{1 - \text{ACR} + \text{OR} \times \text{ACR}}\right|}$$

Como ejemplo, suponga que el odds ratio es OR = 0,73 y el riesgo control supuesto es RCS = 0,3. Entonces el efecto sobre el riesgo es 62 menos por 1000:

$$\text{number fewer per 1000} = 1000 \times \left(0.3 - \frac{0.73 \times 0.3}{1 - 0.3 + 0.73 \times 0.3}\right)$$

$$= 1000 \times \left(0.3 - \frac{0.219}{1 - 0.3 + 0.219}\right) = 1000 \times (0.3 - 0.238) = 61.7$$

El NNT es 17:

$$\text{NNT} = \frac{1}{\left|\left(0.3 - \frac{0.73 \times 0.3}{1 - 0.3 + 0.73 \times 0.3}\right)\right|} = \frac{1}{\left|0.3 - \frac{0.219}{1 - 0.3 + 0.219}\right|} = \frac{1}{|0.3 - 0.238|} = 16.2.$$

12.5.4.4 Cálculo del cociente de riesgos a partir del odds ratio (OR)

Como los cocientes de riesgos son más fáciles de interpretar que los odds ratios, pero los odds ratios tiene propiedades matemáticas favorables, un revisor puede decidir realizar un metanálisis basado en los odds ratios, pero expresar los resultados como un cociente de riesgos resumen (o una reducción del riesgo relativo). Lo anterior requiere de un riesgo control supuesto (RCS). Entonces

$$\mathrm{RR} = \frac{\mathrm{OR}}{1 - \mathrm{ACR} \times (1 - \mathrm{OR})}$$

a menudo será razonable realizar esta transformación mediante la mediana del riesgo del grupo control de los estudios en el metanálisis.

12.5.4.5 Cálculo de los límites de confianza

Los límites de confianza para los RRA y los NNT se pueden calcular al aplicar la fórmula anterior a los límites de confianza inferior y superior para el estadístico resumen (DR, CR u OR) (Altman 1998). Es de señalar que este intervalo de confianza no incorpora la incertidumbre alrededor del riesgo del grupo control (RGC).

En el caso de lo que convencionalmente se considera resultados estadísticamente no significativos (por ejemplo, el intervalo de confianza del OR o el CR incluye el valor 1) uno de los límites de confianza indicará beneficio y el otro daño. Por lo tanto, es necesario el uso apropiado de las palabras “poco” y “más” para cada límite cuando se presentan los resultados en cuanto a los eventos. Para los NNT los dos límites de confianza se deben etiquetar como NNTB y NNTD para indicar la dirección del efecto en cada caso. El intervalo de confianza para el NNT incluirá una “discontinuidad”: dentro del intervalo habrá un NNTB infinitamente grande, el cual cambiará a un NNTD infinitamente grande.

12.6 Interpretación de los resultados continuos (incluidas las diferencias de medias estandarizadas)

12.6.1 Metanálisis con resultados continuos

Cuando los resultados son continuos, los revisores tienen varias opciones para presentar los resultados agrupados. Si todos los estudios han utilizado las mismas unidades, un metanálisis puede generar una estimación agrupada en esas unidades, como una diferencia en la respuesta media (ver por ejemplo la fila de resultados resumen para el edema en el Capítulo 11, Figura 11.5.a). Las unidades de estos resultados pueden ser difíciles de interpretar, particularmente cuando se relacionan con escalas de calificación. Las tablas “Resumen de los hallazgos” deben incluir el mínimo y el máximo de la escala de medición, así como la dirección (nuevamente, ver columna de Edema del Capítulo 11, Figura 11.5.a). Es importante el conocimiento de los cambios más pequeños en la puntuación que perciben los pacientes (la diferencia mínima importante) y que pueden facilitar grandemente la interpretación de los resultados. Conocer la diferencia mínima importante permite a los revisores y a los usuarios colocar los resultados en contexto y los revisores deben señalar la diferencia mínima importante (si se conoce) en la columna Comentarios de su tabla “Resumen de los hallazgos”.

Cuando los estudios han utilizado instrumentos diferentes para medir el mismo constructo, se puede utilizar una diferencia de medias estandarizada (DME) en el metanálisis para combinar los datos continuos (ver Capítulo 9, Sección 9.2.3.2). Para la interpretación clínica dicho análisis puede ser menos útil que la dicotomización de las respuestas y la presentación de las proporciones de los pacientes que se benefician. Hay métodos disponibles para crear datos dicotómicos a partir de las medias y las desviaciones estándar informadas, pero se necesitan suposiciones que es posible que no se cumplan (Suissa 1991, Walter 2001).

La DME expresa el efecto de la intervención en unidades estándar en lugar de en las unidades de medición originales. La DME es la diferencia en los efectos medios en los grupos experimental y control dividida por la desviación estándar agrupada de los resultados de los participantes (ver Capítulo 9, Sección 9.2.3.2). El valor de la DME depende, por lo tanto, del tamaño del efecto (la diferencia entre las medias) y de la desviación estándar de los resultados (la variabilidad inherente entre los participantes).

Sin una guía los médicos y los pacientes tienen muy poca idea de cómo interpretar los resultados presentados como DME. Existen muchas posibilidades para reexpresar dichos resultados de maneras más útiles, las que se muestran a continuación.

12.6.2 Reexpresión de las DME mediante la regla empírica para los tamaños del efecto

Existe una regla empírica para interpretar la DME (o los "tamaños del efecto"), que ha surgido principalmente de investigadores en ciencias sociales. Un ejemplo es el siguiente: 0,2 representa un efecto pequeño, 0,5 un efecto moderado y 0,8 un efecto grande (Cohen 1988). Existen variaciones (por ejemplo, < 0,41 = pequeño; 0,40 a 0,70 = moderado; > 0,70 = grande). Los revisores pudieran considerar incluir una regla empírica en la columna Comentarios de una tabla "Resumen de los hallazgos". Sin embargo, algunos metodólogos creen que estas interpretaciones son problemáticas porque la importancia de un hallazgo para el *paciente* depende del contexto y no se ajusta a planteamientos genéricos.

12.6.3 Reexpresión de las DME mediante la transformación del odds ratio

Existe una transformación disponible de una DME a un odds ratio (logarítmico), basada en la suposición de que la variable continua subyacente tiene una distribución logística con desviación estándar similar en los dos grupos de intervención (Furukawa 1999, Chinn 2000). Es poco probable que la suposición se ajuste exactamente y los resultados se pueden considerar una aproximación. El odds ratio logarítmico se calcula como

$$\ln\mathrm{OR} = \frac{\pi}{\sqrt{3}}\mathrm{SMD},$$

(o aproximadamente 1,81xDME). El odds ratio resultante se puede combinar con un riesgo supuesto en el grupo control para obtener una reducción del riesgo absoluto como en la Sección 12.5.4.3. Este riesgo del grupo control se refiere a la proporción de personas que han tenido alguna mejoría (no especificada) en el resultado continuo ("que respondieron"). La Tabla 12.6.a muestra algunos resultados ilustrativos de este método. Estos NNT se pueden convertir a personas por 1000 mediante la fórmula 1000/NNT.

Tabla 12.6.a: NNT equivalentes para DME específicas para varias "proporción de personas con mejoría" en el grupo control

Proporción de personas con mejoría en el grupo control	10%	20%	30%	40%	50%	60%	70%	80%	90%
DME = 0,1	57	33	26	23	23	24	28	37	66
DME = 0,2	27	16	13	12	12	13	15	20	36
DME = 0,5	9	6	5	5	5	6	7	10	18
DME = 0,8	5	4	3	3	4	4	5	7	14
DME = 1,0	4	3	3	3	3	4	5	7	13

12.6.4 Reexpresión de las DME mediante un instrumento familiar

La posibilidad final para interpretar la DME es expresarla en las unidades de uno o más de los instrumentos de medición específicos. La multiplicación de una DME por una desviación estándar típica entre personas para una escala particular produce una estimación de la diferencia en las puntuaciones medias del resultado (experimental versus control) en dicha

escala. La desviación estándar se pudo obtener como la desviación estándar agrupada de las puntuaciones iniciales en uno de los estudios. Para reflejar mejor la variación entre personas en la práctica, puede ser preferible utilizar una desviación estándar de un estudio observacional representativo. Por lo tanto, el efecto agrupado se reexpresa en las unidades originales de ese instrumento particular y es posible interpretar la relevancia y la repercusión clínicas del efecto de la intervención. Sin embargo, los revisores deben conocer que esta nueva transformación de los tamaños del efecto puede ser errónea si se aplica a estudios individuales en lugar de a una medida resumen del efecto (Scholten 1999). Considere dos estudios que *utilizaron* el mismo instrumento y observaron el mismo efecto, pero observaron diferente variabilidad entre los participantes (quizás debido a diferentes criterios de inclusión). Luego, la nueva transformación mediante las desviaciones estándar diferentes de estos estudios produciría diferentes tamaños del efecto para *la misma escala* y *el mismo efecto*.

12.7 Conclusiones

12.7.1 Sección Conclusiones de una revisión Cochrane

Las conclusiones de los revisores de una revisión Cochrane se dividen en implicaciones para la práctica e implicaciones para la investigación. Al decidir cuáles son estas implicaciones, es útil considerar cuatro factores: la calidad de las pruebas, el equilibrio de beneficios y daños, los valores y preferencias y la utilización de los recursos (Eddy 1990). La consideración de estos factores incluye valoraciones y esfuerzos que van más allá del trabajo de la mayoría de los revisores.

12.7.2 Implicaciones para la práctica

Establecer conclusiones acerca de la utilidad práctica de una intervención implica equilibrar, implícita o explícitamente, las estimaciones de los beneficios y los daños, así como de los costos. Determinar dicho equilibrio y posteriormente hacer recomendaciones para una acción va más allá de una revisión sistemática y requiere información adicional y valoraciones informadas que son típicamente del dominio de quienes desarrollan guías de práctica clínica. Los autores de revisiones Cochrane no deben hacer recomendaciones.

Si los revisores se sienten obligados a establecer acciones que los médicos y pacientes deben realizar, deberían (después de describir la calidad de las pruebas y el equilibrio entre los beneficios y los daños) destacar las diferentes acciones que pudieran ser consistentes con patrones particulares de valores y preferencias. También se deben destacar otros factores que pudieran influir en la decisión, incluido cualquier factor conocido que se esperaría que modificara los efectos de la intervención, el riesgo inicial o el estado del paciente, los costos y quién asumiría estos costos, así como la disponibilidad de recursos. Los revisores deberían asegurar que consideraron todos los resultados importantes para el paciente, incluidos los que tenían pocos datos disponibles. Este proceso implica un alto nivel de explicación acerca de las valoraciones sobre los valores o preferencias vinculadas con los diferentes resultados. El más alto nivel de explicación incluiría un análisis económico formal con un análisis de sensibilidad que incluya diferentes suposiciones sobre los valores o preferencias, lo cual está más allá del alcance de la mayoría de las revisiones Cochrane (aunque las mismas bien pudieran utilizarse para dichos análisis) (Mugford 1989, Mugford 1991). Lo anterior se analiza en el Capítulo 15.

Una revisión sobre el uso de anticoagulantes en pacientes con cáncer para aumentar la supervivencia (Akl 2007) proporciona un ejemplo del establecimiento de implicaciones clínicas para situaciones donde hay compensaciones importantes entre los efectos deseables e indeseables de la intervención. “La decisión de comenzar el tratamiento con heparina en un paciente con cáncer para beneficiar la supervivencia debe equilibrar los beneficios y desventajas e integrar los valores y preferencias del paciente (Haynes 2002). Los pacientes con una alta preferencia por la prolongación de la supervivencia (incluso si esta prolongación

es corta) y una aversión limitada a la hemorragia que no consideran el tratamiento con heparina como una carga pueden optar por el uso de la heparina, mientras es posible que los que tienen aversión a la hemorragia y a la carga relacionada del tratamiento con heparina no lo hagan".

12.7.3 Implicaciones para la investigación

Las conclusiones de la revisión deben ayudar a las personas a tomar decisiones bien informadas acerca de investigaciones futuras en la atención sanitaria. Las "Implicaciones para la investigación" deben comentar la necesidad de investigaciones futuras y la naturaleza de las mismas que sería más recomendable. Se ha propuesto el siguiente formato para el informe de las recomendaciones para la investigación ("EPICOT") (Brown 2006):

- E (Evidencia o pruebas): ¿cuáles son las pruebas actuales?
- P (Población): diagnóstico, estadio de la enfermedad, comorbilidad, factor de riesgo, sexo, edad, grupo étnico, criterios específicos de inclusión y exclusión, ámbito clínico.
- I (Intervención): tipo, frecuencia, dosis, duración, factor pronóstico.
- C (Comparación): placebo, atención habitual, tratamiento alternativo.
- O (Resultado, en inglés "Outcome"): ¿qué resultados clínicos o relacionados con el paciente necesitarán medir, mejorar, influenciar o efectuar los investigadores? ¿qué métodos de medición se deben utilizar?
- T (Tiempo): fecha de la búsqueda bibliográfica o la recomendación.

Otros factores que se pudieran considerar en las recomendaciones incluyen la carga de la enfermedad abordada, los períodos de tiempo (p.ej. duración del seguimiento, duración de la intervención) y el tipo de estudio que sería el mas adecuado en las investigaciones posteriores (Brown 2006).

Los revisores Cochrane deben asegurarse de incluir los aspectos PICO de este formato. También es útil señalar el tipo de estudio, así como cualquier característica del diseño que aborde mejor la pregunta de investigación.

Una revisión de medias compresivas para la prevención de la trombosis venosa profunda en pasajeros de líneas aéreas proporciona un ejemplo donde hay algunas pruebas convincentes de un beneficio de la intervención. "Esta revisión muestra que la pregunta de los efectos sobre la TVP asintomática del uso versus ningún uso de las medias compresivas en el tipo de personas estudiado en estos ensayos clínicos se debe considerar actualmente como respondida. Es posible justificar investigaciones adicionales para investigar los efectos relativos de diferentes intensidades de compresión de las medias o de las medias con otras estrategias preventivas. Los ensayos clínicos aleatorios adicionales para abordar la incertidumbre restante acerca de los efectos del uso versus ningún uso de las medias de compresión sobre resultados como muerte, embolismo pulmonar y TVP deberán ser grandes." (Clarke 2006).

Una revisión del contacto terapéutico para los trastornos de ansiedad proporciona un ejemplo de las implicaciones para la investigación cuando no se han encontrado estudios elegibles. "Esta revisión destaca la necesidad de ensayos clínicos controlados aleatorios para evaluar la efectividad del contacto terapéutico para la reducción de los síntomas de ansiedad en personas con diagnóstico de trastornos de ansiedad. Los ensayos clínicos futuros deben ser rigurosos en cuanto al diseño y la realización, y los informes posteriores deben incluir descripciones de alta calidad de todos los aspectos de la metodología para permitir la evaluación e interpretación de los resultados". (Robinson 2007).

12.7.4 Errores frecuentes al establecer las conclusiones

Un error frecuente cuando no hay pruebas concluyentes es confundir la "falta de pruebas de un efecto" con "las pruebas de ningún efecto". Cuando no hay pruebas concluyentes es incorrecto señalar que se muestra que la intervención "no tiene efecto" o "no es diferente" de la intervención control. Es más seguro informar que los datos, con sus intervalos de confianza, son compatibles con una reducción o un aumento en los resultados. Cuando haya

una tendencia "positiva" pero estadísticamente no significativa, con frecuencia los revisores la describen como "promisoria", mientras que un efecto "negativo" de la misma magnitud con frecuencia no se describe como "signo de alarma". Este lenguaje puede ser perjudicial.

Otro error es enmarcar la conclusión en términos alentadores. Por ejemplo, los revisores pudieran escribir "los estudios incluidos fueron demasiado pequeños para detectar una reducción en la mortalidad" cuando los estudios incluidos mostraron una reducción o incluso un aumento en la mortalidad que no logró alcanzar los niveles convencionales de significación estadística. Una manera de evitar errores como éste es considerar los resultados de forma cegada; es decir, considerar cómo los resultados se presentarían y enmarcarían en las conclusiones si la dirección de los resultados fuera la opuesta. Si el intervalo de confianza para la estimación de la diferencia en los efectos de las intervenciones se superpone al valor nulo, el análisis es compatible con un verdadero efecto beneficioso y con un verdadero efecto perjudicial. Si una de las posibilidades se menciona en la conclusión, también se debe mencionar la otra posibilidad.

Otro error frecuente es establecer conclusiones que van más allá de las pruebas. A menudo se hace de forma implícita, sin hacer referencia a la información o las valoraciones adicionales que se utilizan al establecer las conclusiones acerca de las implicaciones de una revisión para la práctica. Incluso cuando la información adicional y las valoraciones explícitas apoyan las conclusiones acerca de las implicaciones de una revisión para la práctica, los revisores pocas veces realizan revisiones sistemáticas de la información adicional. Además, las implicaciones para la práctica a menudo dependen de circunstancias y valores específicos que se debe tomar en cuenta. Como se ha señalado, los revisores siempre deben ser cautos cuando establecen conclusiones acerca de las implicaciones para la práctica y no deben hacer recomendaciones.

12.8 Información del capítulo

Autores: Holger J Schünemann, Andrew D Oxman, Gunn E Vist, Julian PT Higgins, Jonathan J Deeks, Paul Glasziou y Gordon H Guyatt del Grupo Cochrane de Métodos de Aplicabilidad y Recomendaciones.

La versión en inglés de este capítulo se debe citar como: Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, Guyatt GH. Chapter 12: Interpreting results and drawing conclusions. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: Jonathan Sterne, Michael Borenstein y Rob JM Scholten por sus contribuciones al texto.

Declaración de interés: Holger Schünemann, Andrew Oxman, Gunn Vist, Paul Glasziou y Gordon Guyatt han tenido, en varios grados, funciones principales en el GRADE Working Group del cual han surgido muchas de las ideas en este capítulo.

Recuadro 12.8.a: El Grupo Cochrane de Métodos de Aplicabilidad y Recomendaciones

Se prevé la evolución continua de las metodologías descritas en este capítulo. Las áreas principales en las cuales tendrán lugar discusiones relevantes son el Grupo Cochrane de Métodos de Aplicabilidad y Recomendaciones (GCMAR) y el GRADE Working Group. Ambos grupos de discusión aceptan nuevos participantes con deseos de aprender más y contribuir al desarrollo posterior de la calificación de la calidad de las pruebas y a enmarcar aspectos de la aplicación de las revisiones Cochrane.

El Grupo Cochrane de Métodos de Aplicabilidad y Recomendaciones (GCMAR) está compuesto por individuos con interés y experiencia en la interpretación, aplicabilidad y transferencia de los resultados de las revisiones sistemáticas a individuos y grupos. El objetivo del GCMAR es explorar el proceso de ir de las pruebas a las recomendaciones en la atención sanitaria. El objetivo final es hacer este proceso tan riguroso como sea posible.

Las áreas específicas que actualmente se consideran importantes incluyen:
evaluación de la calidad de las pruebas (www.gradeworkinggroup.org);
variación del efecto con el riesgo inicial;
predicción del beneficio de la tasa o la gravedad del evento esperado en el paciente;
consideración de cómo la fuerza de las pruebas y la magnitud y la precisión de los efectos influyen en las implicaciones; y
consideración de cómo los valores de las personas influyen en las implicaciones cuando se equilibran los beneficios y los daños según las características clínicas individuales.

12.9 Referencias

Aguilar 2005
Aguilar MI, Hart R. Oral anticoagulants for preventing stroke in patients with non-valvular atrial fibrillation and no previous history of stroke or transient ischemic attacks. *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art No: CD001927.

Akl 2007
Akl EA, Kamath G, Kim SY, Yosuico V, Barba M, Terrenato I, Sperati F, Schünemann HJ. Oral anticoagulation for prolonging survival in patients with cancer. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art No: CD006466.

Alonso-Coello 2006
Alonso-Coello P, Zhou Q, Martinez-Zapata MJ, Mills E, Heels-Ansdell D, Johanson JF, Guyatt G. Meta-analysis of flavonoids for the treatment of haemorrhoids. *British Journal of Surgery* 2006; 93: 909-920.

Altman 1998
Altman DG. Confidence intervals for the number needed to treat. *BMJ* 1998; 317: 1309-1312.

Bhandari 2004
Bhandari M, Busse JW, Jackowski D, Montori VM, Schünemann H, Sprague S, Mears D, Schemitsch EH, Heels-Ansdell D, Devereaux PJ. Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials. *Canadian Medical Association Journal* 2004; 170: 477-480.

Brophy 2001
Brophy JM, Joseph L, Rouleau JL. Beta-blockers in congestive heart failure. A Bayesian meta-analysis. *Annals of Internal Medicine* 2001; 134: 550-560.

Brown 2006
Brown P, Brunnhuber K, Chalkidou K, Chalmers I, Clarke M, Fenton M, Forbes C, Glanville J, Hicks NJ, Moody J, Twaddle S, Timimi H, Young P. How to formulate research recommendations. *BMJ* 2006; 333: 804-806.

Cates 1999
Cates C. Confidence intervals for the number needed to treat: Pooling numbers needed to treat may not be reliable. *BMJ* 1999; 318: 1764-1765.

Chambers 2005
Chambers BR, Donnan GA. Carotid endarterectomy for asymptomatic carotid stenosis. *Cochrane Database of Systematic Reviews* 2005, Issue 4. Art No: CD001923.

Chinn 2000
Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine* 2000; 19: 3127-3131.

Cina 2000
Cina CS, Clase CM, Haynes RB. Carotid endarterectomy for symptomatic carotid stenosis. *Cochrane Database of Systematic Reviews* 2000, Issue 2. Art No: CD001081.

Clarke 2006
Clarke M, Hopewell S, Juszczak E, Eisinga A, Kjeldstrøm M. Compression stockings for preventing deep vein thrombosis in airline passengers. *Cochrane Database of Systematic Reviews* 2006, Issue 2. Art No: CD004002.

Cohen 1988
Cohen J. *Statistical Power Analysis in the Behavioral Sciences* (2nd edition). Hillsdale (NJ): Lawrence Erlbaum Associates, Inc., 1988.

Dans 2007
Dans AM, Dans L, Oxman AD, Robinson V, Acuin J, Tugwell P, Dennis R, Kang D. Assessing equity in clinical practice guidelines. *Journal of Clinical Epidemiology* 2007; 60: 540-546.

Devereaux 2004
Devereaux PJ, Choi PT, El-Dika S, Bhandari M, Montori VM, Schünemann HJ, Garg AX, Busse JW, Heels-Ansdell D, Ghali WA, Manns BJ, Guyatt GH. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. *Journal of Clinical Epidemiology* 2004; 57: 1232-1236.

Eddy 1990
Eddy DM. Clinical decision making: from theory to practice. Anatomy of a decision. *JAMA* 1990; 263: 441-443.

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

Furukawa 1999
Furukawa TA. From effect size into number needed to treat. *The Lancet* 1999; 353: 1680.

Gibson 2007
Gibson JN, Waddell G. Surgical interventions for lumbar disc prolapse. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art No: CD001350.

GRADE Working Group 2004
GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328: 1490-1494.

Guyatt 1994
Guyatt GH, Sackett DL, Cook DJ. 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? Evidence-Based Medicine Working Group. *JAMA* 1994; 271: 59-63.

Guyatt 2006a
Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B, Raskob G, Lewis SZ, Schünemann H. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American College of Chest Physicians Task Force. *Chest* 2006; 129: 174-181.

Guyatt 2006b
Guyatt G, Vist G, Falck-Ytter Y, Kunz R, Magrini N, Schünemann H. An emerging consensus on grading recommendations? *ACP Journal Club* 2006; 144: A8-A9.

Guyatt 2008a
Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ. What is 'quality of evidence' and why is it important to clinicians? *BMJ* 2008; 336: 995-998.

Guyatt 2008b
Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; 336: 924-926.

Hawe 2004
Hawe P, Shiell A, Riley T, Gold L. Methods for exploring implementation variation and local context within a cluster randomised community intervention trial. *Journal of Epidemiology and Community Health* 2004; 58: 788-793.

Haynes 2002
Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *ACP Journal Club* 2002; 136: A11-A14.

Hoffrage 2000
Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Medicine. Communicating statistical information. *Science* 2000; 290: 2261-2262.

Levine 2004
Levine MN, Raskob G, Beyth RJ, Kearon C, Schulman S. Hemorrhagic complications of anticoagulant treatment: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest* 2004; 126: 287S-310S.

Lumley 2004
Lumley J, Oliver SS, Chamberlain C, Oakley L. Interventions for promoting smoking cessation during pregnancy. *Cochrane Database of Systematic Reviews* 2004, Issue 4. Art No: CD001055.

McIntosh 2005
McIntosh HM, Jones KL. Chloroquine or amodiaquine combined with sulfadoxine-pyrimethamine for treating uncomplicated malaria. *Cochrane Database of Systematic Reviews* 2005, Issue 4. Art No: CD000386.

McQuay 1997
McQuay HJ, Moore A. Using numerical results from systematic reviews in clinical practice. *Annals of Internal Medicine* 1997; 126: 712-720.

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: 1003-1006.

Mugford 1991
Mugford M, Piercy J, Chalmers I. Cost implications of different approaches to the prevention of respiratory distress syndrome. *Archives of Disease in Childhood* 1991; 66: 757-764.

Oxman 2002
Oxman A, Guyatt G. When to believe a subgroup analysis. In: Guyatt G, Rennie D (editors). *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chicago (IL): AMA Press, 2002.

Resnicow 1993
Resnicow K, Cross D, Wynder E. The Know Your Body program: a review of evaluation studies. *Bulletin of the New York Academy of Medicine* 1993; 70: 188-207.

Robinson 2007
Robinson J, Biley FC, Dolk H. Therapeutic touch for anxiety disorders. *Cochrane Database of Systematic Reviews* 2007, Issue 3. Art No: CD006240.

Sackett 2000
Sackett DL, Richardson WS, Rosenberg W, Haynes BR. *Evidence-Based Medicine: How to Practice and Teach EBM* (2nd edition). Edinburgh (UK): Churchill Livingstone, 2000.

Salpeter 2007
Salpeter S, Greyber E, Pasternak G, Salpeter E. Risk of fatal and nonfatal lactic acidosis with metformin use in type 2 diabetes mellitus. *Cochrane Database of Systematic Reviews* 2007, Issue 4. Art No: CD002967.

Scholten 1999
Scholten RJPM. From effect size into number needed to treat [letter]. *The Lancet* 1999; 453: 598.

Schünemann 2006a
Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 13. Applicability, transferability and adaptation. *Health Research Policy and Systems* 2006; 4: 25.

Schünemann 2006b
Schünemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A, Fahy BF, Gould MK, Horan KL, Krishnan JA, Manthous CA, Maurer JR, McNicholas WT, Oxman AD, Rubenfeld G, Turino GM, Guyatt G. An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *American Journal of Respiratory and Critical Care Medicine* 2006; 174: 605-614.

Smeeth 1999
Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses - sometimes informative, usually misleading. *BMJ* 1999; 318: 1548-1551.

Suissa 1991
Suissa S. Binary methods for continuous outcomes: a parametric alternative. *Journal of Clinical Epidemiology* 1991; 44: 241-248.

Thompson 2000
Thompson DC, Rivara FP, Thompson R. Helmets for preventing head and facial injuries in bicyclists. *Cochrane Database of Systematic Reviews* 2000, Issue 2. Art No: CD001855.

Walter 2001
Walter SD. Number needed to treat (NNT): estimation of a measure of clinical benefit. *Statistics in Medicine* 2001; 20: 3947-3962.

PARTE 3: TEMAS ESPECIALES

ÍNDICE

Capítulo 13: Inclusión de estudios no aleatorizados

Capítulo 14: Efectos adversos

Capítulo 15: Incorporación de pruebas económicas

Capítulo 16: Temas especiales en estadística

Capítulo 17: Resultados notificados por el paciente

Capítulo 18: Revisiones de datos de pacientes individuales

Capítulo 19: Metanálisis prospectivo

Capítulo 20: Investigación cualitativa y revisiones Cochrane

Capítulo 21: Revisiones en salud pública y promoción de salud

Capítulo 22: Resúmenes de revisiones

ÍNDICE

CAPÍTULO 13: INCLUSIÓN DE ESTUDIOS NO ALEATORIZADOS

PUNTOS CLAVE

13.1 INTRODUCCIÓN

13.1.1 De qué trata este capítulo

Recuadro 13.1.a: Algunos tipos de diseños de ENA utilizados para evaluar los efectos de intervenciones

13.1.2 ¿Por qué considerar estudios no aleatorizados?
13.1.3 Aspectos clave acerca de la inclusión de estudios no aleatorizados en una revisión Cochrane
13.1.4 La importancia de un protocolo para una revisión Cochrane que incluya estudios no aleatorizados
13.1.5 Estructura de las secciones posteriores en el capítulo

13.2 DESARROLLO DE LOS CRITERIOS PARA INCLUIR ESTUDIOS NO ALEATORIZADOS

13.2.1 ¿Qué es diferente cuando se incluyen estudios no aleatorizados?
13.2.1.1 Inclusión de estudios aleatorizados y no aleatorizados
13.2.1.2 Evaluación de los beneficios y los daños
13.2.1.3 Determinar qué tipo de estudios no aleatorizados incluir
13.2.1.4 Diferenciar entre preguntas de investigación etiológicas y de efectividad
13.2.2 Recomendaciones y recursos disponibles para apoyar a los revisores
Tabla 13.2.a: Lista de características del diseño del estudio (estudios con asignación a las intervenciones a nivel individual)
Tabla 13.2.b: Lista de características del diseño del estudio (estudios con asignación a las intervenciones a nivel de grupo)
13.2.3 Resumen

13.3 BÚSQUEDA DE ESTUDIOS NO ALEATORIZADOS

13.3.1 ¿Qué es diferente cuando se incluyen estudios no aleatorizados?
13.3.1.1 Exhaustividad de la estrategia de búsqueda
13.3.1.2 Identificación de los ENA en las búsquedas
13.3.1.3 Revisión de citas y resúmenes
13.3.2 Recomendaciones y recursos disponibles para apoyar a los revisores
13.3.3 Resumen

13.4 SELECCIÓN DE LOS ESTUDIOS Y OBTENCIÓN DE LOS DATOS

13.4.1 ¿Qué es diferente cuando se incluyen estudios no aleatorizados?
13.4.2 Recomendaciones y recursos disponibles para apoyar a los revisores
Recuadro 13.4.a: Guía del usuario para la obtención de los datos/evaluación del estudio mediante la lista de verificación en la Tabla 13.2.a o la Tabla 13.2.b
13.4.3 Resumen

13.5 EVALUACIÓN DEL RIESGO DE SESGO EN LOS ESTUDIOS NO ALEATORIZADOS

13.5.1 ¿Qué es diferente cuando se incluyen estudios no aleatorizados?
13.5.1.1 Fuentes de sesgo en los estudios no aleatorizados
13.5.1.2 Evidencia de riesgo de sesgo en los estudios no aleatorizados
13.5.2 Recomendaciones y recursos disponibles para apoyar a los revisores

13.5.2.1 Consideraciones generales al evaluar el riesgo de sesgo en los estudios no aleatorizados
13.5.2.2 Factores de confusión y ajuste
13.5.2.3 Herramientas para evaluar la calidad metodológica o el riesgo de sesgo en los estudios no aleatorizados
13.5.2.4 Limitaciones prácticas al evaluar el riesgo de sesgo en los estudios no aleatorizados

13.5.3 Resumen

13.6 SÍNTESIS DE LOS DATOS DE LOS ESTUDIOS NO ALEATORIZADOS

13.6.1 ¿Qué es diferente cuando se incluyen estudios no aleatorizados?
13.6.2 Recomendaciones y recursos disponibles para apoyar a los revisores
13.6.2.1 Control por factores de confusión
13.6.2.2 Combinación de los estudios
13.6.2.3 Análisis de la heterogeneidad
13.6.2.4 Cuando se considera que la combinación no es apropiada
13.6.3 Resumen

13.7 INTERPRETACIÓN Y DISCUSIÓN

13.7.1 Retos en la interpretación de las revisiones Cochrane de efectividad que incluyen estudios no aleatorizados

Introducción

13.7.1.1 ¿Se han incluido todos los estudios importantes y relevantes?
13.7.1.2 ¿Se ha evaluado adecuadamente el riesgo de sesgo de los estudios incluidos?
13.7.2 Evaluación de la fuerza de la evidencia proporcionada por revisiones que incluyen estudios no aleatorizados
13.7.3 Recomendaciones para los posibles revisores

13.8 INFORMACIÓN DEL CAPÍTULO

Recuadro 13.8.a: El Grupo Cochrane de Métodos de Estudios No Aleatorizados

13.9 REFERENCIAS

CAPÍTULO 13: INCLUSIÓN DE ESTUDIOS NO ALEATORIZADOS

Autores: Barnaby C Reeves, Jonathan J Deeks, Julian PT Higgins y George A Wells del Grupo Cochrane de Métodos de Estudios No Aleatorizados.

Puntos clave

- Para algunas revisiones Cochrane no es posible responder la pregunta de interés mediante ensayos aleatorizados y puede justificarse que los revisores incluyan estudios no aleatorizados.
- Es probable que los sesgos potenciales sean mayores para los estudios no aleatorizados comparados con los ensayos aleatorizados, por lo que los resultados siempre se deberían interpretar con precaución cuando se incluye este tipo de estudios en las revisiones y los metanálisis. Pueden surgir dudas concretas con respecto a las diferencias entre las personas en los diferentes grupos de intervención (sesgo de selección) y en los estudios que no informan explícitamente que hayan sido precedidos de un protocolo (sesgo de notificación).
- Se recomienda que los criterios de elegibilidad, la obtención de los datos y la evaluación crítica de los estudios incluidos hagan énfasis en aspectos específicos del diseño del estudio (p.ej. qué partes del estudio se diseñaron de forma prospectiva) en lugar de "calificar" el diseño del estudio (como casos y controles versus cohortes).
- El riesgo de sesgo en los estudios no aleatorizados se puede evaluar de una manera similar a la utilizada para los ensayos aleatorizados, aunque se debe poner más atención a la posibilidad de sesgo de selección.
- Los metanálisis de los estudios no aleatorizados deberían considerar cómo se abordaron los posibles factores de confusión, y considerar la probabilidad del aumento de la heterogeneidad a partir de los factores de confusión residuales y de otros sesgos que varían entre los estudios.

13.1 Introducción

13.1.1 De qué trata este capítulo

Este capítulo ha sido preparado por el Grupo de Métodos de Estudios No Aleatorizados (GMENA) de la Colaboración Cochrane (ver Recuadro 13.8.a). Su objetivo es apoyar a los revisores que se plantean incluir estudios no aleatorizados en revisiones Cochrane. Los estudios no aleatorizados (ENA) se definen aquí como cualquier estudio cuantitativo que calcule la efectividad de una intervención (daños o beneficios) que no utilice la asignación al azar para asignar las unidades a grupos de comparación. Se incluyen los estudios en los que la asignación ocurre en el curso de las decisiones habituales de tratamiento o por la decisión de las personas, es decir, los estudios habitualmente llamados *observacionales*. Existen muchos tipos de estudios de intervención no aleatorizados, incluidos los estudios de cohortes, los estudios de casos y controles, los estudios controlados tipo antes y después, los estudios de series de tiempo interrumpido y los ensayos controlados que utilizan estrategias inapropiadas de asignación al azar (en algunas ocasiones llamados estudios cuasialeatorizados). El Recuadro 13.1.a resume algunos términos de diseños de estudios utilizados con frecuencia para los estudios no aleatorizados. En la Sección 13.5.1 se explica por qué no se aconseja que se utilicen necesariamente estas calificaciones en las revisiones Cochrane.

Este capítulo tiene como objetivo describir los retos particulares que surgen si se incluyen ENA en una revisión Cochrane, a través de consideraciones teóricas o epidemiológicas, investigaciones empíricas y discusiones entre los miembros del GMENA. Este capítulo hace recomendaciones acerca de qué hacer cuando es posible apoyar las recomendaciones sobre la base de las pruebas o la teoría establecida. Cuando no es posible hacer recomendaciones, el capítulo intenta establecer los pro y los contra de las acciones alternativas e identificar preguntas para investigaciones metodológicas adicionales.

Los revisores que se plantean incluir ENA en una revisión Cochrane no deberían comenzar con este capítulo a menos que ya estén familiarizados con el proceso de preparación de una revisión sistemática de ensayos aleatorizados. El formato y los pasos básicos de una revisión Cochrane deberían ser los mismos si se incluyen sólo ensayos aleatorizados o se incluyen ENA. Se remite al lector a la Parte 2 del Manual para una descripción detallada de estos pasos. Cada paso que se da para realizar una revisión sistemática es más difícil cuando se incluyen ENA y el revisor debe tratar de incluir epidemiólogos y metodólogos expertos en el equipo de revisión. Como ejemplo de esta colaboración una revisión de ENA incluyó nueve revisores, cinco de los cuales eran metodólogos (Siegfried 2003).

Recuadro 13.1.a: Algunos tipos de diseños de ENA utilizados para evaluar los efectos de intervenciones

Los diseños se diferencian a continuación por calificaciones de uso común y las descripciones son intencionalmente no específicas porque las calificaciones se interpretan de manera diferente con respecto a los detalles. El GMENA no recomienda el uso de estas calificaciones por los motivos explicados en la Sección 13.5.1.

Ensayo controlado no aleatorizado.	Un estudio experimental en el cual las personas se asignan a diferentes intervenciones mediante métodos que no son aleatorizados.
Estudio controlado tipo antes y después.	Un estudio en el cual las observaciones se hacen antes y después de la implementación de una intervención en un grupo que recibe la intervención y en un grupo control que no la recibe.
Estudio de series de tiempo interrumpido.	Un estudio que utiliza observaciones en múltiples puntos temporales antes y después de una intervención (la "interrupción"). El diseño intenta detectar si la intervención ha tenido un efecto significativamente mayor que cualquier tendencia subyacente en el tiempo.
Estudio con controles históricos.	Un estudio que compara un grupo de participantes que recibe una intervención con un grupo similar del pasado que no la recibió.
Estudios de cohortes.	Un estudio en el cual un grupo definido de personas (la cohorte) es seguido en el tiempo para examinar las asociaciones entre las diferentes intervenciones recibidas y los resultados posteriores. Un estudio de cohorte "prospectivo" recluta los participantes antes de cualquier intervención y los sigue hacia el futuro. Un estudio de cohorte "retrospectivo" identifica los sujetos de registros del pasado que describen las intervenciones recibidas y los sigue desde el momento de dichos registros.
Estudio de casos y controles.	Un estudio que compara las personas con un resultado de interés específico ("casos") con personas de las misma población fuente pero sin ese resultado ("controles"), para examinar la asociación entre el resultado y la exposición previa (p.ej. recibir una intervención). Este diseño es particularmente útil cuando el resultado es poco frecuente.
Estudio transversal.	Un estudio que recopila información de intervenciones (pasadas o actuales) y resultados actuales de salud, es decir, limitados al estadio de salud, para un grupo de personas en un punto particular en el tiempo, para examinar las asociaciones entre los resultados y la exposición a las intervenciones.
Series de casos (estudio longitudinal no controlado).	Las observaciones se hacen en una serie de individuos, habitualmente todos reciben la misma intervención, antes y después de una intervención pero sin un grupo control.

13.1.2 ¿Por qué considerar estudios no aleatorizados?

La colaboración Cochrane se centra particularmente en revisiones sistemáticas de ensayos aleatorizados porque es más probable que proporcionen información no sesgada que otros diseños de estudios acerca de los efectos diferenciales de formas alternativas de atención sanitaria. Las revisiones de ENA sólo es probable que se realicen cuando no es posible responder la pregunta de interés mediante una revisión de ensayos aleatorizados. El GMENA considera que puede justificarse que los revisores incluyan ENA moderadamente susceptibles al sesgo. En general, el GMENA considera que hay tres motivos fundamentales para incluir ENA en una revisión Cochrane:

a) Examinar el caso de realizar un ensayo aleatorizado al proporcionar una evaluación explícita de la debilidad de los ENA disponibles. Los hallazgos de una revisión de ENA también pueden ser útiles para informar el diseño de un ensayo aleatorizado posterior, p.ej. mediante la identificación de subgrupos relevantes.
b) Proporcionar pruebas de los efectos (beneficiosos o perjudiciales) de intervenciones que no pueden tener una asignación aleatoria, o que es muy poco probable que se estudien en ensayos aleatorizados. En estos contextos puede ser útil una revisión desinteresada (libre de sesgo y parcialidad) que informe de manera sistemática los hallazgos y las limitaciones de los ENA.
c) Proporcionar pruebas de los efectos (beneficiosos o perjudiciales) que no se pueden estudiar en ensayos aleatorizados, como resultados a largo plazo y poco frecuentes, o resultados que no se conocía que eran importantes cuando se realizaron ensayos aleatorizados grandes.

Otros tres motivos se citan con frecuencia para apoyar las revisiones sistemáticas de ENA pero son justificaciones insuficientes:

d) Estudiar los efectos en grupos de pacientes no reclutados en ensayos aleatorizados (como niños, mujeres embarazadas, personas de edad avanzada). Aunque es importante considerar si los resultados de los ensayos se pueden generalizar a las personas que se excluyen de los mismos, no está claro que se pueda lograr cuando se consideran estudios no aleatorizados. Independientemente de si las estimaciones de los ENA coinciden o no con las de los ensayos aleatorizados, siempre existe la posibilidad de sesgo en los resultados de los ENA, y que se establezcan conclusiones erróneas.
e) Complementar las pruebas de ensayos aleatorizados existentes. Agregar no pruebas aleatorias a aleatorias puede cambiar una estimación imprecisa pero no sesgada en una estimación precisa pero sesgada, es decir, cambiar de una incertidumbre no conveniente a un error inaceptable.
f) Cuando el efecto de una intervención es muy grande. Implícitamente, esta es una justificación a partir del resultado o *post hoc*, ya que es necesario realizar la revisión (o alguna síntesis de las pruebas) para observar el tamaño probable de los efectos. Aunque es más fácil argumentar que es menos probable que los efectos grandes se puedan explicar completamente por la presencia de sesgo que los efectos pequeños (Glasziou 2007), para la práctica de la atención sanitaria aún es importante obtener estimaciones no sesgadas de la magnitud de los efectos grandes para tomar decisiones clínicas y económicas (Reeves 2006). De forma que todavía se necesitan ensayos aleatorizados para los efectos grandes (y no necesitan ser grandes si los efectos son realmente grandes). Puede haber una oposición ética a los ensayos aleatorizados de intervenciones que ya se sospecha que se asocian con grandes beneficios como resultado de una revisión sistemática de ENA, lo que hace difícil asignar al azar a los participantes, y las intervenciones que se postula tienen un efecto grande pueden ser difíciles de asignar al azar por otros motivos, (p.ej. cirugía versus no cirugía). Sin embargo, la justificación para una revisión sistemática de ENA en estas circunstancias se debe clasificar como (b), es decir, intervenciones que es poco probable que se asignen al azar, en lugar de (f).

13.1.3 Aspectos clave acerca de la inclusión de estudios no aleatorizados en una revisión Cochrane

Los ensayos aleatorizados son el diseño preferido para estudiar los efectos de las intervenciones en la atención sanitaria porque en la mayoría de los casos el ensayo aleatorizado es el diseño de estudio que es menos probable que esté sesgado. Cualquier revisión Cochrane debe considerar el riesgo de sesgo en los estudios primarios individuales, incluidas la probable dirección y magnitud del sesgo (ver Capítulo 8). Una revisión que incluya ENA también necesita revisores que valoren este aspecto. El principio de considerar el riesgo de sesgo es exactamente el mismo. Sin embargo, es probable que los sesgos potenciales sean mayores para los ENA comparados con los ensayos aleatorizados. Los revisores deberían considerar (a) la debilidad de los diseños que se han utilizado (por ejemplo, al señalar su capacidad para determinar causalidad), (b) la realización de los estudios mediante una evaluación cuidadosa de su riesgo de sesgo, especialmente, (c) el potencial para el sesgo de selección y los factores de confusión de los cuales todos los ENA son sospechosos y (d) el potencial para el sesgo de notificación, incluido la notificación selectiva de resultados.

La susceptibilidad al sesgo de selección (que en este *Manual* significa las diferencias en las características iniciales de los individuos en diferentes grupos de intervención, en lugar de si la muestra seleccionada es representativa de la población) se considera ampliamente como la diferencia principal entre los ensayos aleatorizados y los ENA. La asignación al azar con una ocultación adecuada de la secuencia de asignación reduce la posibilidad de sesgo de selección sistemático en los ensayos aleatorizados, de manera que las diferencias en las características entre los grupos se pueden atribuir al azar. En los ENA, la asignación a los grupos depende de otros factores, a menudo desconocidos. Los factores de confusión aparecen cuando el sesgo de selección da lugar a desequilibrios entre los grupos intervención y control (o en los grupos de casos y controles en los estudios de casos y controles) en cuanto a los factores pronósticos, es decir, las distribuciones de los factores difieren entre los grupos *y* los factores se asocian con el resultado. Los factores de confusión pueden tener dos efectos en un metanálisis: (a) cambiar la estimación del efecto de la intervención (sesgo sistemático) y (b) aumentar la variabilidad de los efectos observados e introducir una heterogeneidad excesiva entre los estudios (Deeks 2003). Es importante considerar ambos efectos posibles (ver Sección 13.6.1). La Sección 13.5 proporciona una discusión más detallada de la susceptibilidad al sesgo en los ENA.

13.1.4 La importancia de un protocolo para una revisión Cochrane que incluya estudios no aleatorizados

El Capítulo 2 establece la importancia de redactar un protocolo para una revisión Cochrane antes de realizar la revisión. Como las decisiones metodológicas tomadas durante una revisión de ENA son complejas y pueden afectar los hallazgos de la revisión, un protocolo para una revisión que incluya ENA es aún más importante. La justificación para hacer una revisión que incluya ENA (ver Sección 13.1.2) se debe documentar en el protocolo. El protocolo debe incluir muchos más detalles que una revisión de ensayos aleatorizados y preespecificar las decisiones metodológicas clave acerca de los métodos a utilizar y los análisis planificados. El protocolo debe especificar detalles que no son relevantes para los ensayos aleatorizados (p.ej. los métodos planificados para identificar los factores de confusión potenciales y para evaluar la susceptibilidad de los estudios primarios a los factores de confusión), así como proporcionar más detalles acerca de los pasos estándar en el proceso de revisión que son más difíciles cuando se incluyen ENA (p.ej especificación de los criterios de elegibilidad y la estrategia de búsqueda para identificar los estudios elegibles).

El GMENA reconoce que a veces no es posible preespecificar todas las decisiones acerca de los métodos utilizados en una revisión. Sin embargo, los revisores deberían tener como objetivo

tomar todas las decisiones acerca de los métodos para la revisión sin referirse a los hallazgos de los estudios primarios e informar acerca de las decisiones metodológicas que tuvieron que tomar o modificar después de la obtención de los datos acerca de los hallazgos de los estudios.

13.1.5 Estructura de las secciones posteriores en el capítulo

Cada una de las secciones de este capítulo, que se focaliza en mostrar los diferentes pasos del proceso de revisión, está estructurada de la misma manera. Primero, para un determinado estadio, se resume qué es diferente (en comparación con los ensayos aleatorizados) cuando se incluyen ENA en una revisión Cochrane y, de ser aplicable, se describen los aspectos conceptuales que es necesario considerar. La primera parte incluye las pruebas relevantes, donde existen. Segundo, se resumen las recomendaciones y, de estar disponibles, se describen los recursos existentes que están disponibles para apoyar a los revisores.

13.2 Desarrollo de los criterios para incluir estudios no aleatorizados

13.2.1 ¿Qué es diferente cuando se incluyen estudios no aleatorizados?

13.2.1.1 Inclusión de estudios aleatorizados y no aleatorizados

Los revisores pueden desear incluir ENA en una revisión porque sólo es posible identificar un escaso número de ensayos aleatorizados, o debido a las limitaciones percibidas en los ensayos aleatorizados. En este capítulo se recomienda firmemente que los revisores no intenten combinar las pruebas de los ensayos aleatorizados y los ENA. Esta recomendación significa que los criterios para los diseños de los estudios a incluir deberían especificar generalmente estudios aleatorizados o no aleatorizados cuando se intenta evaluar el efecto de una intervención sobre un resultado particular. Sin embargo, una revisión única pudiera consistir de revisiones “componentes” que incluyan diferentes diseños de estudios para diferentes resultados, por ejemplo, ensayos aleatorizados para evaluar los beneficios y ENA para evaluar los daños (ver Capítulo 14). Alternativamente, cuando son convenientes las pruebas provenientes de ensayos aleatorizados pero es poco probable que estén disponibles, los criterios de elegibilidad pudieran estar razonablemente estructurados para decir que sólo se incluirán ENA cuando se encuentre que no hay ensayos aleatorizados disponibles. En su momento, cuando esta revisión se actualice, los ENA podrán ser retirados si hay ensayos aleatorizados disponibles. Cuando existan estudios aleatorizados y no aleatorizados de una intervención y, por alguno de los motivos expuestos en la Sección 13.1.2 se incluyan ambos en la revisión, los mismos se deberían presentar por separado; alternativamente, si hay un número adecuado de ensayos aleatorizados, se pueden incluir comentarios de ENA relevantes en la sección Discusión de una revisión, aunque pocas veces es particularmente útil.

13.2.1.2 Evaluación de los beneficios y los daños

Las revisiones Cochrane tienen como objetivo cuantificar los efectos beneficiosos y perjudiciales, esperados e inesperados, de las intervenciones en la atención sanitaria. La mayoría de las revisiones hace una estimación de los beneficios esperados de una intervención que se evalúa en ensayos aleatorizados. Los ensayos aleatorizados pueden informar algunos de los daños de una intervención, ya sean los esperados, para cuya evaluación se diseñó el ensayo, o los que no se esperaban pero que se obtuvieron en el ensayo como parte de la monitorización estándar de la seguridad. Sin embargo, muchos daños graves de una intervención son demasiado poco frecuentes o no aparecen durante el período de seguimiento de los ensayos aleatorizados, por lo que no se informarán. Por lo

tanto, una de las funciones más importantes de las revisiones de ENA es evaluar los daños potenciales no esperados o poco frecuentes de las intervenciones (motivo [c] en la Sección 13.1.1). Es difícil establecer los criterios para seleccionar estudios importantes y relevantes para evaluar efectos adversos e inesperados poco frecuentes o a largo plazo. Aunque las fortalezas y las debilidades relativas de los diferentes diseños de estudios son las mismas para los resultados beneficiosos, la selección del diseño del estudio a incluir puede depender de la frecuencia y la importancia de un resultado. Por ejemplo, para algunos resultados adversos poco frecuentes es posible que sólo estén disponibles estudios de series de casos o de casos y controles. Los diseños de los estudios que son más susceptibles al sesgo pueden ser aceptables para evaluar eventos graves a falta de mejores pruebas.

Los factores de confusión pueden ser una amenaza de menor importancia para la validez de una revisión cuando se investigan efectos perjudiciales, inesperados o poco frecuentes de las intervenciones, que cuando se investigan efectos esperados, ya que se ha señalado que el "factor de confusión por indicación" influye principalmente sobre las decisiones de tratamiento con respecto a los resultados acerca de los cuales los médicos tienen dudas fundamentales. Sin embargo, los factores de confusión nunca se pueden descartar porque las mismas características de ser factores de confusión para los efectos esperados también pueden ser factores de confusión directos para los efectos inesperados, o estar correlacionados con características que son factores de confusión.

Un aspecto relacionado es la necesidad de diferenciar entre *cuantificar* y *detectar* el efecto de una intervención. La cuantificación de los beneficios pretendidos de una intervención (maximizar la precisión de la estimación y minimizar la susceptibilidad al sesgo) es fundamental cuando se sopesan los méritos relativos de las intervenciones alternativas para la misma afección. Una revisión debe tratar de cuantificar los daños de una intervención y minimizar la susceptibilidad al sesgo tanto como sea posible. Sin embargo, si una revisión puede mostrar más allá de cualquier duda razonable que una intervención provoca un daño particular, es posible que la precisión y la susceptibilidad al sesgo de la estimación del efecto no sean fundamentales. En otras palabras, la gravedad del daño puede sobrepasar cualquier beneficio de la intervención. Esta situación es más probable que ocurra cuando hay intervenciones opuestas para una afección.

13.2.1.3 Determinar qué tipo de estudios no aleatorizados incluir

Un ensayo aleatorizado es un diseño de estudio prospectivo experimental que incluye específicamente la asignación al azar de los participantes a las intervenciones. Aunque hay variaciones en el diseño de los ensayos aleatorizados (que incluyen la asignación al azar de individuos, por conglomerados o grupos [*clusters*] o partes del cuerpo; ensayos con múltiples brazos, ensayos factoriales y ensayos cruzados), constituyen una categoría de estudio característica. Por el contrario, los ENA cubren varios diseños fundamentalmente diferentes, muchos de los cuales se concibieron originalmente en el contexto de la epidemiología etiológica. Algunos de ellos se resumen en el Recuadro 13.1.a, aunque esta lista no es exhaustiva, y muchos estudios combinan ideas de diferentes diseños básicos. Como se discute en la Sección 13.2.2, estas calificaciones no se aplican de manera consistente. La diversidad de diseños de ENA hace que surjan dos interrogantes. Primero, ¿se deberían incluir en una revisión todos los diseños de ENA de una pregunta particular de efectividad? Segundo, si los revisores no incluyen todos los diseños de ENA, ¿qué criterios se deberían utilizar para decidir qué diseños de estudios incluir y cuáles excluir?

En general se acepta que se deberían establecer criterios para limitar el tipo de pruebas incluidas en una revisión sistemática. El motivo principal es que el riesgo de sesgo varía entre los estudios. Por este motivo, muchas revisiones Cochrane sólo incluyen ensayos aleatorizados (cuando están disponibles). Por el mismo motivo se ha señalado que los revisores sólo deberían incluir ENA que sea menos probable que estén sesgados. No es útil incluir estudios primarios en una revisión cuando es probable que los resultados de los estudios estén sesgados, incluso si no hay mejores pruebas. Lo anterior se debe a que la

estimación errónea del efecto puede ser más perjudicial para los pacientes futuros que la falta de estimación, particularmente si las personas que utilizan las pruebas para tomar decisiones no están al tanto de sus limitaciones (Doll 1993, Peto 1995).

No hay coincidencia acerca de los criterios de diseño de los estudios que se deberían utilizar para limitar la inclusión de ENA en una revisión Cochrane. Una estrategia es incluir solamente los diseños de estudios que proporcionarán estimaciones del efecto razonablemente válidas. Otra estrategia es incluir los mejores diseños de estudios disponibles que se han utilizado para responder una pregunta. La primera estrategia significaría que las revisiones son consistentes e incluyen el mismo tipo de ENA, pero algunas revisiones no incluyen estudios. La segunda estrategia da lugar a diferentes revisiones que incluyen diferentes diseños de estudios según lo que esté disponible. Por ejemplo, pudiera ser completamente apropiado utilizar diferentes criterios para la inclusión cuando se revisan los daños, comparados con los beneficios, de una intervención. Este enfoque ya es evidente en la *Base de Datos Cochrane de Revisiones Sistemáticas* (*Cochrane Database of Systematic Reviews, CDSR*), en la que los editores de algunos Grupos Cochrane de Revisión (GCR) limitan las revisiones a ensayos aleatorizados solamente y otros editores de GCR permiten que se incluyan tipos específicos de ENA en las revisiones (generalmente en áreas de la atención sanitaria en las que son poco frecuentes los ensayos aleatorizados).

Cualquiera que sea el punto de vista que se adopte, los criterios sólo se pueden seleccionar con respecto a la jerarquía de los diseños de los estudios primarios, clasificados en orden de riesgo de sesgo según las características del diseño del estudio. Las “jerarquías de las pruebas” existentes para los estudios de efectividad (Eccles 1996, National Health ans Medical Research Council 1999, Oxford Centre for Evidence-based Medicine 2001) parecen haber surgido en su mayoría al aplicar las jerarquías para preguntas de investigación etiológica a preguntas de efectividad. Por ejemplo, convencionalmente se considera que los estudios de cohortes proporcionan mejores pruebas que los estudios de casos y controles. No está claro que lo anterior siempre sea apropiado porque las jerarquías etiológicas ponen más énfasis en establecer la causalidad (p.ej. relación dosis-respuesta, exposición que precede al resultado) que en la cuantificación válida del tamaño del efecto. Además, los diseños de los estudios utilizados para examinar los efectos de las intervenciones pueden ser mucho más diversos y complejos (Shadish 2002) y es posible que no sean fácilmente asimilables en las jerarquías de pruebas existentes (por ejemplo, ver el orden de los diseños en el Recuadro 13.1.a). Diferentes diseños son susceptibles a diferentes sesgos, y a menudo no está claro qué sesgos tienen el mayor impacto y cómo varían en las situaciones clínicas.

13.2.1.4 Diferenciar entre preguntas de investigación etiológicas y de efectividad

La inclusión de ENA en una revisión Cochrane permite, en principio, la inclusión de estudios verdaderamente observacionales en los que el uso de una intervención ha ocurrido en el curso de la atención sanitaria habitual o de la vida diaria. Para las intervenciones que no están limitadas al ámbito médico lo anterior puede referirse a las intervenciones que el participante de un estudio decide tomar, p.ej. preparados sin prescripción. La inclusión de estudios observacionales en una revisión también permite estudiar exposiciones que no son claramente “intervenciones”, p.ej. decisiones nutricionales, y otras conductas que pueden afectar la salud. Lo anterior introduce un “área gris” entre las pruebas acerca de la efectividad y la etiología. Es importante diferenciar cuidadosamente entre las preguntas de investigación etiológicas y de efectividad relacionadas con una exposición particular. Por ejemplo, los nutricionistas pueden estar interesados en los efectos relacionados con la salud de una dieta que incluye un mínimo de cinco porciones de frutas o vegetales por día (“cinco veces al día”), una pregunta etiológica. Por otra parte, los profesionales de la salud pública pueden estar interesados en los efectos relacionados con la salud de intervenciones que promueven un cambio en la dieta para incluir la dieta “cinco veces al día”, una pregunta de efectividad. Debido a otras diferencias entre los estudios relevantes para estos dos tipos de preguntas (p.ej, duración del seguimiento y resultados investigados), los estudios que abordan el primer tipo de pregunta a menudo se perciben como “mejores” o “más relevantes”, sin reconocer o

tener en cuenta que abordan diferentes preguntas de investigación. En otros casos la intervención sanitaria que se evalúa en el ENA se habrá realizado con un fin que no es mejorar la salud. Por ejemplo, una revisión sobre la circuncisión para prevenir la transmisión del VIH incluyó ENA en los que la circuncisión se había realizado por motivos religiosos y culturales (Siegfried 2003), y no está claro si el uso de la intervención para fines sanitarios hubiera tenido el mismo efecto.

13.2.2 Recomendaciones y recursos disponibles para apoyar a los revisores

Los revisores deberían verificar primero con los editores del GCR en el cual proponen registrar su protocolo si existe una política específica establecida por el GCR acerca de la inclusión de ENA en la revisión. Los revisores también deberían discutir con los editores el grado de asesoramiento disponible en el GCR, ya que es probable que necesiten más apoyo que para una revisión que incluya solamente ensayos aleatorizados, e intentar reclutar metodólogos informados para su equipo de revisión. Desafortunadamente, el GMENA no está actualmente en disposición de colaborar con los autores de revisiones particulares, pero estimula a los revisores que incluyen ENA en sus revisiones que expongan sus experiencias al GMENA, particularmente cuando su experiencia apoye o contradiga las experiencias descritas en este capítulo.

Los revisores que planifiquen revisar los efectos adversos (daños) de una intervención deberían leer el Capítulo 14, preparado por el Grupo de Métodos de Efectos Adversos.

Se recomienda que los revisores utilicen las características explícitas del diseño del estudio (nótese bien: no las calificaciones de diseño del estudio) cuando decidan qué tipo de ENA incluir en una revisión. Los miembros del GMENA han desarrollado dos listas que se pueden utilizar para estos fines, aunque la experiencia con su uso es limitada. La Tabla 13.2.a y la Tabla 13.2.b describen listas separadas para los estudios con asignación individual y por conglomerados. Se han agrupado 16 (ó 15) ítems bajo cuatro encabezados:

1. ¿Hubo una comparación?
2. ¿Cómo se crearon los grupos?
3. ¿Qué partes del estudio fueron prospectivas?
4. ¿En qué variables se evaluó la comparabilidad (entre los grupos que recibieron diferentes intervenciones)?

Los ítems están diseñados para caracterizar aspectos clave de los estudios que, sobre la base de las experiencias de los miembros del GMENA y los “principios fundamentales” (en lugar de las pruebas), se sospecha que definen las principales categorías del diseño de los estudios o se asocian con susceptibilidad al sesgo. Las tablas indican qué aspectos se asocian con diferentes diseños de ENA, identificados mediante calificaciones que son más específicas que las del Recuadro 13.1.a. No hay consenso total acerca del uso de estas calificaciones (columnas). Estos desacuerdos no significan que los ítems en las filas no sean apropiados o que estén descritos de forma deficiente; el valor de las listas depende del acuerdo entre los revisores cuando clasifican los estudios primarios. También se hará la propuesta de que estas listas se utilicen como listas de verificación en el proceso de obtención de los datos y como parte de la evaluación crítica de los estudios (Sección 13.4.2 y Sección 13.5.2). Las instrucciones para el uso de los ítems como listas de verificación en el Recuadro 13.4.a proporcionan explicación adicional de los ítems.

Varias organizaciones realizan actualmente revisiones sistemáticas de ENA en las que no hay o hay muy pocos ensayos aleatorizados. A menudo las revisiones son realizadas por organizaciones responsables de emitir políticas o recomendaciones a los profesionales sanitarios, p.ej. el *National Institute for Health and Clinical Excellence* (NICE), la *Canadian Agency for Drugs and Technologies in Health* (CADTH), y las realizan equipos de revisores sistemáticos en departamentos universitarios de ciencias de la salud. En general, los revisores

de estos equipos intentan aplicar a los ENA los métodos desarrollados para las revisiones sistemáticas de ensayos aleatorizados. Los grupos incluyen:

- Grupo Cochrane para una Práctica y Organización Sanitaria Efectivas (EPOC) (*Effective Practice and Organisation of Care*) (www.epoc.cochrane.org).
- *The Centre for Reviews and Dissemination* (www.york.ac.uk/inst/crd).
- Centro EPPI, Institute of Education, Universidad de Londres (eppi.ioe.ac.uk).
- *The Effective Public Health Practice Project* (EPHPP), Canadian Ministry of Health, Long-Term Care and the City of Hamilton, Public Health Services (enlace a la lista de revisiones del EPHPP: old.hamilton.ca/phcs/ephpp).

Los GCR y los revisores Cochrane tienden a limitar la inclusión de ENA debido al diseño del estudio o la calidad metodológica, y reconocen que el diseño de los ENA influye en la susceptibilidad al sesgo. Por ejemplo, el GCR EPOC acepta protocolos que incluyan estudios de series de tiempo interrumpido y estudios controlados tipo antes y después, pero no otros diseños de ENA. Otras revisiones han limitado la inclusión a los estudios con “calidad metodológica adecuada” (Taggart 2001).

Tabla 13.2.a: Lista de características del diseño del estudio (estudios con asignación a las intervenciones a nivel individual)

	ECA	Cuasi ECA	ECNA	ECAD	ECP	ECR	ECH	ECCS	CC	ET	CAD	IC/SC
¿Hubo una comparación:												
entre dos o más grupos de participantes que recibieron intervenciones diferentes?	S	S	S	S	S	S	S	S	S	S	N	N
dentro del mismo grupo de participantes en el tiempo?	P	P	N	S	N	N	N	N	N	N	S	N
¿Los participantes se asignaron a los grupos mediante:												
asignación al azar oculta?	S	N	N	N	N	N	N	N	N	N	na	na
asignación al azar cuasialeatoria?	N	S	N	N	N	N	N	N	N	N	na	na
otra acción de los investigadores?	N	N	S	P	N	N	N	N	N	N	na	na
diferencias en el tiempo?	N	N	N	N	N	N	S	N	N	N	na	na
diferencias de ubicación?	N	N	P	P	P	P	P	na	na	na	na	na
decisiones de tratamiento?	N	N	N	P	P	P	N	N	N	P	na	na
preferencias de los participantes?	N	N	N	P	P	P	N	N	N	P	na	na
según el resultado?	N	N	N	N	N	N	N	S	S	P	na	na
algún otro proceso? (especificar)												
¿Qué partes del estudio fueron prospectivas:												
identificación de los participantes?	S	S	S	P	S	N	P*	S	N	N	P	P
evaluación inicial y asignación a la intervención?	S	S	S	P	S	N	P*	S	N	N	na	na
evaluación de los resultados?	S	S	S	P	S	P	P	S	N	N	P	P
generación de hipótesis?	S	S	S	S	S	S	S	S	P	P	P	na
¿En qué variables se evaluó la comparabilidad entre los grupos:												
posibles factores de confusión?	P	P	P	P	P	P	P	P	P	P	N	na
evaluación inicial de las variables de resultado?	P	P	P	S	P	P	P	N	N	N	N	na

S=Sí; P=Posiblemente; P*=Posible para un grupo solamente; N=No; na=no aplicable. Nota: Nótese que "posiblemente" se utiliza en la tabla para indicar celdas en las que puede ser el caso "S" o "N". No se debería utilizar como una opción de respuesta cuando se aplica la lista de verificación; en caso de incertidumbre, la respuesta debería ser "no se puede valorar" (ver Recuadro 13.4.a).
ECA=Ensayo controlado aleatorizado; CuasiECA=Ensayo controlado cuasialeatorizado; ECNA=Ensayo controlado no aleatorizado; ECAD=estudio controlado tipo antes y después; ECP=estudio de cohortes prospectivas; ECR=estudio de cohortes retrospectivas; ECH=Ensayo de controles históricos; ECCS=estudio de casos y controles secundario; CC= estudio de casos y controles; ET=estudio transversal; CAD=comparaciones antes y después; IC/SC=informe de casos/series de casos.

Tabla 13.2.b: Lista de características del diseño del estudio (estudios con asignación a las intervenciones a nivel de grupo)

	ECAC	Cuasi ECAC	ECNAC	STIC	ECCAD	STI	ECAD	ETEco
¿Hubo una comparación:								
entre dos o más grupos de conglomerados que recibieron intervenciones diferentes?	S	S	S	S	S	S	S	S
dentro del mismo grupo de conglomerados en el tiempo?	P	P	N	S	N	N	N	N
¿Los conglomerados se asignaron a los grupos mediante:								
asignación al azar oculta?	S	N	N	N	N	N	N	N
asignación al azar cuasialeatoria?	N	S	N	N	N	N	N	N
otra acción de los investigadores?	N	N	S	P	P	N	N	N
diferencias en el tiempo?	N	N	N	S	S	S	S	N
diferencias de ubicación?	N	N	P	P	P	N	N	P
decisiones de política/salud pública?	Na	na	P	P	P	P	na	na
preferencias del conglomerado?	Na	na	P	P	P	P	na	na
algún otro proceso? (especificar)								
¿Qué partes del estudio fueron prospectivas:								
identificación de los conglomerados participantes?	S	S	S	P	P	P	P	N
evaluación inicial y asignación a la intervención?	S	S	S	P	P	P	P	N
evaluación de los resultados?	S	S	S	P	P	P	P	N
generación de hipótesis?	S	S	S	S	S	S	S	P
¿En qué variables se evaluó la variabilidad entre los grupos:								
posibles factores de confusión?	P	P	P	P	P	P	P	P
evaluación inicial de las variables de resultado?	P	P	P	S	S	S	S	N

Nótese que "conglomerado" se refiere a una entidad (p.ej. una organización), no necesariamente a un grupo de participantes; "grupo" se refiere a uno o más conglomerados; ver Recuadro 13.4.a.

Nótese que "posiblemente" se utiliza en la tabla para indicar las celdas en las que *cualquiera* puede ser el caso, es decir "S" o "N". No se debería utilizar como una opción de respuesta cuando se aplica la lista de verificación; en caso de incertidumbre, la respuesta debería ser "no se puede valorar" (vern Recuadro 13.4.a)

S=Sí; P=Posiblemente; P*=Posible para un grupo solamente; N=No; NR=No se requiere. ECAC=ensayo controlado aleatorizado por conglomerado; CuasiECAC =Ensayo controlado cuasialeatorizado por conglomerado; ECNAC=ensayo controlado no aleatorizado por conglomerado; STIC=Series de tiempo interrumpido controladas (Shadisch 2002); ECCAD=estudio controlado de cohortes tipo antes y después (Shadisch 2002); STI= Serie de tiempo interrumpido; ECAD= estudio de cohortes tipo antes y después (Shadisch 2002); ETEco=Estudio transversal ecológico.

13.2.3 Resumen

- Los revisores deberían justificar cuidadosamente sus motivos para la inclusión de ENA en su revisión sistemática.
- Los revisores deberían consultar la política editorial del GCR en el cual ellos se proponen registrar su protocolo, con respecto a la inclusión de ENA. Los revisores deberían considerar el grado de asesoría metodológica disponible en el GCR y el apoyo metodológico que tienen en su equipo.
- Los revisores deberían especificar los criterios de elegibilidad según lo que los investigadores hicieron (es decir, aspectos importantes del diseño del estudio), así como los factores relacionados con la pregunta de interés específica para la revisión (es decir, intervención, población, problema de salud) para evitar la ambigüedad. Se recomienda que para hacerlo los revisores utilicen los ítems de la lista de verificación del GMENA o una lista similar.
- Los revisores también necesitan información acerca de lo que hicieron los investigadores en los estudios primarios para categorizar los estudios identificados. Se recomienda que los revisores utilicen las listas de características del diseño del estudio del GMENA o una herramienta similar para estos fines, y registren los aspectos importantes del diseño del estudio que no sean claros o no se describan.
- Los revisores que examinen preguntas acerca de los efectos adversos (daños) de las intervenciones deberían leer el Capítulo 14.

13.3 Búsqueda de estudios no aleatorizados

13.3.1 ¿Qué es diferente cuando se incluyen estudios no aleatorizados?

13.3.1.1 Exhaustividad de la estrategia de búsqueda

Cuando una revisión se propone incluir ensayos aleatorizados solamente, un principio clave de la búsqueda de estudios elegibles es que los revisores deberían tratar tanto como sea posible de identificar todos los ensayos aleatorizados sobre la pregunta de revisión que se han iniciado. Por lo tanto, se les recomienda a los revisores que busquen en los registros de ensayos, resúmenes de congresos, literatura gris, etc., así como en bases de datos bibliográficas estándar como MEDLINE, PUBMED, EMBASE (ver Capítulo 6). Se ha señalado que para una revisión sistemática es necesario realizar una búsqueda exhaustiva para evitar los sesgos de publicación. Es fácil argumentar que los autores de revisiones que incluyen ENA deberían hacer lo mismo (Petticrew 2011). Sin embargo, es importante establecer las premisas que sustentan la justificación original para realizar una búsqueda exhaustiva y considerar con mucho cuidado si se aplican a las revisiones de ENA. Las premisas son:

a) Existe una población finita de ensayos aleatorizados que investiga la pregunta de revisión.
b) Todos los ensayos aleatorizados en esta población se pueden identificar a través de una búsqueda suficientemente exhaustiva porque los ensayos aleatorizados se identifican con relativa facilidad, hay registros disponibles y son difíciles de realizar sin financiación y aprobación ética, lo que crea una “pista de la intervención” (Chan 2004).
c) Todos los ensayos aleatorizados en esta población, si se realizan bien, proporcionan información valiosa.
d) La facilidad de acceso a la información acerca de estos ensayos aleatorizados se relaciona con sus hallazgos, de manera que los ensayos identificados con más facilidad

pueden ser un subgrupo sesgado. Este es el sesgo de publicación: es más probable que los estudios con hallazgos favorables y estadísticamente significativos se publiquen en lugares accesibles (ver Capítulo 10, Sección 10.2). Debido a que es menos probable que los estudios más pequeños produzcan estos hallazgos, la imposibilidad de identificar todos los estudios puede dar lugar a una asimetría en el gráfico de embudo. En teoría es posible obtener una respuesta no sesgada al identificar todos los ensayos aleatorizados, es decir, mediante una búsqueda exhaustiva para descubrir los estudios pequeños, no significativos o no favorables. Los estudios más pequeños también pueden presentar diferencialmente otros sesgos, lo que representa una causa alternativa de asimetría en el gráfico de embudo. Los riesgos de estos sesgos se comprenden razonablemente bien y se pueden evaluar (Capítulo 10, Sección 10.4).

No está claro que estas premisas se apliquen de la misma manera a los ENA.

La Sección 13.2.1.3 señala que los ENA incluyen diferentes diseños y que es difícil categorizarlos. Incluso si los revisores son capaces de establecer criterios específicos del diseño de los estudios con respecto a los cuales se deban evaluar para su inclusión los ENA potenciales, muchos de los ENA potencialmente elegibles no proporcionarán información suficiente que permita clasificarlos.

Existe un problema adicional en la definición exacta cuando aparece un ENA. Por ejemplo, ¿es un ENA elegible un estudio de cohortes que ha recopilado datos sobre las intervenciones y resultados de interés, pero que no ha examinado su asociación? ¿Es un ENA elegible un resultado producido por un ordenador en un fichero que incluye un *odds ratio* calculado para la asociación relevante? Por lo tanto, es difícil definir “una población finita de ENA” para una pregunta de revisión concreta. Es posible que no se puedan rastrear en absoluto algunos ENA realizados, es decir, no se encontrarán siquiera en la proverbial “gaveta del fondo”.

A pesar de los problemas para definir lo que constituye un ENA elegible, la identificación real de los ENA representa retos importantes. Lo anterior no tiene que ver solamente con la descripción deficiente, sino también con:

- la falta de registros de ENA;
- la indexación deficiente de características importantes del diseño, etc.;
- los ENA no siempre necesitan de aprobación ética (al menos en el pasado);
- los ENA no siempre tienen un patrocinador o un financiador de la investigación; y
- los ENA no son siempre bien ejecutados según un protocolo preespecificado.

No hay pruebas de que los sesgos de notificación afecten a los ensayos aleatorizados y los ENA de manera diferente. Sin embargo, es difícil creer que los sesgos de notificación puedan afectar a los ENA *menos* que a los ensayos aleatorizados, debido al creciente número de aspectos asociados con la realización y la publicación de los ensayos aleatorizados que tratan de prevenir los sesgos de notificación, los cuales con frecuencia están ausentes en los ENA (protocolo preespecificado, aprobación ética que incluya el progreso y las publicaciones finales, la declaración CONSORT (Moher 2001), los registros de ensayos y la indexación del tipo de publicación en las bases de datos bibliográficas). A diferencia de la situación con los ensayos aleatorizados, no se conocen la probable magnitud y los determinantes del sesgo de publicación.

No están claros los beneficios de la búsqueda exhaustiva para los ENA y éste es un tema que requiere investigación adicional. Es posible que los estudios más difíciles de encontrar sean los más sesgados, si la dificultad para identificarlos se relaciona con el diseño y el tamaño de muestra deficientes. Con las revisiones de los ensayos aleatorizados, la búsqueda exhaustiva ofrece una protección potencial contra el sesgo porque existe una población definida de estudios elegibles, por lo que los estudios pequeños con hallazgos no significativos deberían finalmente ser identificados. Con las revisiones de ENA, incluso si se puede definir una población *teóricamente* finita de estudios elegibles, no se tiene una

confianza similar en que se puedan identificar estudios faltantes con hallazgos no significativos.

13.3.1.2 Identificación de los ENA en las búsquedas

Es fácil diseñar una estrategia de búsqueda que identifique toda la evidencia acerca de una intervención al crear cadenas de búsqueda para las características de la población y la enfermedad, la intervención y posiblemente el comparador. Cuando una revisión pretende incluir ensayos aleatorizados solamente, hay varios enfoques disponibles para limitar la estrategia de búsqueda a éstos (ver Capítulo 6):

a) Búsqueda de revisiones previas de la pregunta de revisión.
b) Uso de recursos como CENTRAL o los registros específicos de los GCR que son "ricos" en ensayos aleatorizados.
c) Uso de filtros metodológicos y campos de indexación como el tipo de publicación en MEDLINE para limitar las búsquedas a estudios que es probable que sean ensayos aleatorizados.
d) Búsqueda en registros de ensayos.

Limitar la búsqueda a diseños de estudios no aleatorizados particulares es más difícil. De los enfoques anteriores sólo (a) y (b) es probable que sean de alguna ayuda. Los revisores deberían ciertamente buscar en los registros específicos de los GCR los ENA potencialmente relevantes. Algunos GCR (p.ej. el grupo EPOC) incluyen tipos particulares de ENA en los registros específicos del GCR (los revisores deberían verificarlo con sus GCR). El proceso de identificación de los estudios para la inclusión en CENTRAL significa que algunos, pero no todos, los ENA están incluidos, por lo que las búsquedas en esta base de datos no serán completamente exhaustivas, incluso para los estudios que utilizan un diseño particular. No hay bases de datos de ENA similares a CENTRAL.

Como se discute en la Sección 13.2.1.3, los revisores no utilizan de manera consistente las calificaciones del diseño de los estudios y las mismas no se indexan de forma confiable en las bases de datos bibliográficas. Es poco probable que la estrategia (c) sea útil porque las calificaciones del diseño de los estudios que no sean ensayos aleatorizados no se indexan de forma fidedigna en las bases de datos bibliográficas y a menudo los autores de los estudios primarios las utilizan de manera inconsistente. Algunos revisores han tratado de desarrollar y "validar" estrategias de búsqueda para los ENA (Wieland 2005, Fraser 2006, Furlan 2006). Algunos revisores también han intentado optimizar las estrategias de búsqueda para los efectos adversos (ver Capítulo 14, Sección 14.5) (Golder 2006b, Golder 2006c). Debido a que por su naturaleza las revisiones que incluyen ENA consumen mucho tiempo, los intentos por desarrollar estrategias de búsqueda para los ENA no han investigado grandes números de preguntas de revisión. Por lo tanto, los revisores deberían ser cautos al suponer que las estrategias previas se pueden aplicar necesariamente a nuevos temas.

13.3.1.3 Revisión de citas y resúmenes

Habitualmente los ensayos aleatorizados se pueden identificar en los resultados de la búsqueda simplemente a partir de los títulos y los resúmenes, particularmente desde la implementación de los estándares de publicación. Desafortunadamente, los detalles del diseño de los ENA que se requieren para evaluar la elegibilidad a menudo no se describen en los títulos ni en los resúmenes y es necesario acceder al texto completo del estudio.

13.3.2 Recomendaciones y recursos disponibles para apoyar a los revisores

El GMENA no recomienda limitar las estrategias de búsquedas a términos indexados para el diseño del estudio. Sin embargo, se recomienda que los revisores establezcan contacto

con investigadores que hayan tenido éxito en el desarrollo de estrategias de búsqueda eficientes para ENA (ver Sección 13.3.1) y otros revisores que hayan realizado revisiones Cochrane (u otras revisiones sistemáticas) de ENA para examinar preguntas similares a la suya.

Cuando se realizan búsquedas de ENA se recomienda a los revisores que busquen estudios que investiguen todos los efectos de una intervención y no limiten las estrategias de búsqueda a resultados específicos (Capítulo 6). Cuando se realicen búsquedas de ENA con resultados específicos poco frecuentes o a largo plazo (habitualmente adversos o no esperados) de una intervención, puede estar justificada la inclusión de texto libre y términos MeSH para resultados específicos en la estrategia de búsqueda. Los miembros de Grupo de Métodos de Efectos Adversos tiene experiencia con este procedimiento (ver Capítulo 14, Sección 14.5).

Los revisores deberían verificar con los editores de su GCR si el registro específico del GCR incluye estudios con características particulares del diseño de los estudios y deberían solicitar asesoramiento de expertos en recuperación de información en el GCR y en el Grupo de Métodos de Recuperación de Información (ver Capítulo 6, Recuadro 6.7.a).

13.3.3 Resumen

- Para identificar estudios de los efectos beneficiosos esperados de las intervenciones, las estrategias de búsqueda deberían incluir cadenas de búsqueda para la intervención y la población y el problema de salud de interés. Actualmente, no existe métodos recomendados para limitar las estrategias de búsqueda según el diseño del estudio.
- Los revisores que realicen búsquedas de evidencia relacionadas con efectos adversos “sospechados” pueden considerar la búsqueda de resultados específicos de interés (es decir, efectos adversos). Es evidente que este enfoque no se puede utilizar para búsquedas más generales de los posibles efectos adversos de una intervención (ver Capítulo 14, Sección 14.5).
- Es posible que la búsqueda exhaustiva, que se recomienda para los ensayos aleatorizados, no se justifique cuando se examinan ENA. Sin embargo, hasta el presente no hay investigaciones que guíen a los revisores en este importante aspecto.

13.4 Selección de los estudios y obtención de los datos

13.4.1 ¿Qué es diferente cuando se incluyen estudios no aleatorizados?

Los resultados de la búsqueda a menudo contienen una gran cantidad de citas y resúmenes no relevantes que con frecuencia no proporcionan detalles adecuados acerca del diseño de los ENA (que es probable que se necesiten para evaluar la elegibilidad). Por lo tanto, al contrario de la situación en la que se revisan ensayos aleatorizados, puede ser necesario obtener y leer muchas publicaciones completas de estudios para seleccionar los estudios elegibles.

Los revisores necesitan obtener todos los datos requeridos para una revisión sistemática de ensayos aleatorizados (ver Capítulo 7), así como los datos que describen (a) las características del diseño de un estudio primario (ver Sección 13.2.2), (b) los factores de confusión considerados y los métodos utilizados para controlar por dichos factores (ver Sección 13.1.3), (c) los aspectos del riesgo de sesgo específicos para los ENA (ver Sección 13.5.1) y (d) los resultados (ver Sección 13.6.1).

Habitualmente los revisores recopilan información “cruda” acerca de los resultados cuando revisan ensayos aleatorizados, p.ej. para un resultado dicotómico, el número total de

participantes y el número que presentó el resultado en cada grupo. Si los participantes se asignan al azar a los grupos se supone que una comparación de estos datos crudos no está sesgada. Para un ENA, una comparación de los mismos datos crudos "no está ajustada" y es sensible a los factores de confusión. Habitualmente los revisores también describen una comparación "ajustada" calculada a partir de un modelo de regresión que no se puede resumir de la misma manera. Los revisores deberían aún registrar el tamaño de muestra reclutado en cada grupo, el número analizado y el número de eventos, pero también necesitan documentar cualquier estimación ajustada del efecto y sus errores estándar o intervalos de confianza. Estos datos se pueden utilizar para mostrar las estimaciones del efecto ajustadas, así como su precisión en los gráficos de bosque y, de ser apropiado, combinar los datos entre los estudios.

De manera anecdótica, la experiencia de los revisores es que los ENA se informan de manera deficiente por lo que la información requerida es difícil de encontrar, y revisores diferentes pueden recopilar información diferente del mismo artículo. Puede ser necesario adaptar los formularios de obtención de datos a la pregunta de investigación que se estudia. Debido a la diversidad de los estudios potencialmente elegibles y de las formas en las que se describen los mismos, el desarrollo de formularios de obtención de datos puede necesitar numerosas repeticiones en el curso de una revisión de una muestra de estudios primarios. Es casi imposible completar su diseño por adelantado.

Los resultados en los ENA se pueden presentar mediante diferentes medidas del efecto y la incertidumbre de la significación estadística depende del estilo de la publicación y de los análisis realizados. El asesoramiento estadístico por expertos puede ayudar a los revisores a transformar o "trabajar hacia atrás" a partir de la información proporcionada en un artículo hasta obtener una medida del efecto consistente entre los estudios. Las hojas de extracción de datos deberían ser capaces de manejar los diferentes tipos de información acerca de los hallazgos de los estudios que los revisores pueden encontrar.

13.4.2 Recomendaciones y recursos disponibles para apoyar a los revisores

A la vez que proporcionan información para decidir acerca de la elegibilidad, las preguntas de la Tabla 13.2.a y la Tabla 13.2.b representan una lista de verificación conveniente para obtener los datos relevantes de los ENA acerca de las características del diseño del estudio. Al utilizar estas listas de verificación para obtener información acerca de los estudios y decidir sobre la elegibilidad, la intención debe ser documentar lo que los investigadores hicieron en los estudios primarios, en lugar de cómo denominaron los investigadores a sus estudios o qué consideraron que hicieron. Los ítems se deberían registrar como "Sí", "No" o "No se puede valorar". El Recuadro 13.4.a proporciona recomendaciones para utilizar estas tablas como listas de verificación.

Se han desarrollado formularios de obtención de datos para su uso en talleres del GMENA, con el fin de ilustrar la extracción de datos de los ENA. Los mismos incluyen: la lista de verificación del diseño del estudio, plantillas para obtener información acerca de los factores de confusión, su comparabilidad inicial, los métodos utilizados para ajustar por los factores de confusión y las estimaciones del efecto. Estos recursos (disponibles en el sitio web de recursos del *Manua:* www.cochrane.org/resources/handbook) se pueden utilizar como una guía para los tipos de formularios de obtención de datos que los revisores necesitarán. Sin embargo, los revisores deberán adaptar los formularios cuidadosamente para la pregunta de revisión estudiada.

Recuadro 13.4.a: Guía del usuario para la obtención de los datos/evaluación del estudio mediante la lista de verificación en laTabla 13.2.a o la Tabla 13.2.b

Nota: los usuarios deberían tener muy clara la forma en la que se utilizan los términos "grupo" y "conglomerado" en estas tablas. La Tabla 13.2.a sólo se refiere a los grupos, que se utiliza en su sentido convencional y significa un número de participantes individuales. Con la excepción de asignación según el resultado, "grupo" se puede interpretar como sinónimo de "grupo de intervención". La Tabla 13.2.b se refiere a conglomerados y a grupos. En esta tabla "conglomerados" son generalmente las entidades organizacionales como la práctica de medicina familiar o las áreas administrativas, no los individuos. Como en la Tabla 13.2.a, "grupo" es sinónimo de "grupo de intervención" y se utiliza para describir un grupo de unidades asignadas, pero en la Tabla 13.2.b estas unidades son conglomerados en lugar de individuos. Además, aunque los individuos se reúnen en conglomerados, un conglomerado no representa necesariamente un grupo fijo de individuos. Por ejemplo, en estudios asignados por conglomerado, los conglomerados se estudian frecuentemente en dos o más puntos temporales (períodos) en los que diferentes grupos de individuos contribuyen con datos obtenidos en cada punto temporal.

¿Hubo una comparación?
Generalmente los investigadores comparan dos o más grupos que reciben intervenciones diferentes y los grupos se pueden estudiar durante el mismo período de tiempo o en diferentes períodos de tiempo (ver más adelante). En ocasiones los investigadores comparan los resultados en un grupo solamente pero en dos puntos temporales. También es posible que los investigadores hagan ambas cosas, es decir, estudien dos o más grupos y midan los resultados en más de un punto temporal.

¿Los participantes/conglomerados se asignaron a los grupos por?
Estos ítems tienen como objetivo describir cómo se formaron los grupos. No se aplicarán si los estudios no comparan dos o más grupos de sujetos. A menudo la información no se proporciona o es difícil de encontrar en un artículo. Los ítems proporcionados cubren las formas principales en las cuales se pueden formar los grupos. Se puede aplicar más de una opción a un estudio único, aunque algunas opciones son mutuamente excluyentes (es decir, un estudio es aleatorizado o no).

Asignación al azar: la asignación se realizó según una secuencia verdaderamente aleatoria. Estos estudios se incluyen en las recomendaciones estándar en cualquier sección de este *Manual*. Verifique cuidadosamente si la asignación se ocultó adecuadamente hasta que los sujetos se reclutaron definitivamente.

Asignación cuasialeatoria: la asignación se realizó según una secuencia pseudoaleatoria, p.ej. número par/impar de historia clínica o fecha de nacimiento, alternancia. Nota: cuando se utilizan estos métodos el problema es que pocas veces la asignación se ocultó. Estos estudios se han incluido a menudo en revisiones sistemáticas que sólo incluyen ensayos aleatorizados, en los que se utiliza la evaluación del riesgo de sesgo para distinguirlos de los ensayos con asignación al azar apropiada.

Por otra acción de los investigadores: ésta es una categoría en la que se incluye todo y se deberían señalar los detalles adicionales si los investigadores los describen. La asignación sucedió como resultado de alguna decisión o sistema aplicados por los investigadores. Por ejemplo, los sujetos tratados en "unidades" particulares de administración (p.ej. salas hospitalarias, consultas de atención primaria) se "seleccionaron" para recibir la intervención y los sujetos tratados en otras unidades se seleccionaron para recibir la intervención

control.

Diferencias temporales: el reclutamiento a los grupos no ocurrió de forma contemporánea. Por ejemplo, en un estudio de controles históricos los sujetos del grupo control se reclutan generalmente antes en el tiempo que los sujetos del grupo intervención; luego se presenta la intervención y se reclutan los sujetos que la reciben. Generalmente ambos grupos se reclutan en el mismo ámbito. Si el diseño estaba bajo el control de los investigadores se deberían marcar esta opción y "otra acción de los investigadores" para un estudio único. Si el diseño "apareció" por la introducción de una nueva intervención, esta opción y "decisiones de tratamiento" se deberían marcar para un estudio único.

Diferencias de ubicación: se compararon dos o más grupos en diferentes áreas geográficas y la decisión de qué área/s recibieron la intervención y las intervenciones control no se hizo de forma aleatoria. Por lo tanto, esta opción y "otra acción de los investigadores" se deberían marcar para un estudio único.

Decisiones de tratamiento: los grupos de intervención y control se formaron a partir de las variaciones que ocurrieron de forma natural en las decisiones de tratamiento. Esta opción está diseñada para reflejar decisiones de tratamiento tomadas principalmente por los médicos a cargo; la siguiente opción está diseñada para reflejar las decisiones de tratamiento tomadas principalmente según las preferencias de los sujetos. Si las preferencias de tratamiento son uniformes para una "unidad" proveedora determinada o cambia con el tiempo, se deberían marcar esta opción y "diferencias de ubicación" o "diferencias temporales".

Preferencias del paciente: los grupos intervención y control se formaron a partir de las variaciones que ocurrieron de forma natural en las preferencias de los pacientes. Esta opción está diseñada para reflejar decisiones de tratamiento tomadas principalmente según las preferencias de los sujetos; la opción previa está diseñada para reflejar las decisiones de tratamiento tomadas principalmente por los médicos a cargo.

Según el resultado: un grupo de personas que presentó un determinado resultado de interés se comparó con un grupo de personas que no lo presentó, es decir, un estudio de casos y controles. Nota: se debe marcar esta opción para artículos que describen análisis de *múltiples factores de riesgo para un determinado resultado* en una serie grande de sujetos, es decir, cuando la población total del estudio se divide en los que presentaron el resultados y los que no lo presentaron. Estos estudios están mucho más cerca de los estudios de casos y controles secundarios que los estudios de cohortes, incluso cuando se obtienen de forma prospectiva datos longitudinales de pacientes consecutivos.

Opciones adicionales para los estudios por conglomerado

Diferencias de ubicación: ver anteriormente.

Decisiones políticas/salud pública: los grupos de intervención y control se formaron por las decisiones tomadas por personas a cargo de implementar políticas acerca de la salud pública o de la provisión de servicios. Cuando estas decisiones coinciden con conglomerados, o cuando estas personas son los propios investigadores, este ítem se superpone con "otras decisiones de los investigadores" o "preferencias de los conglomerados".

Preferencias de los conglomerados: los grupos intervención y control se formaron a partir de las variaciones que ocurrieron de forma natural en las preferencias de los conglomerados, p.ej. las preferencias colectivas o individuales a nivel de la entidad del conglomerado.

¿Qué partes del estudio fueron prospectivas?

Estos ítems tienen como objetivo describir qué partes del estudio se realizaron de forma prospectiva. En un ensayo controlado aleatorizado estos cuatro ítems serían prospectivos. Para los ENA también es posible que los cuatro sean

prospectivos, aunque es probable que no haya detalles adecuados para discernirlo, particularmente con respecto a la generación de hipótesis. En algunos estudios de cohortes es posible identificar que los participantes se asignaron al tratamiento de forma retrospectiva pero los resultados se evaluaron de forma prospectiva.

¿En qué variables se evaluó la variabilidad de los grupos?
Estas preguntas deberían identificar a los estudios del tipo "antes-después". La evaluación inicial de las variables de resultado es particularmente útil cuando los resultados se miden con escalas continuas, p.ej. el estado de salud o la calidad de vida.

Opciones de respuesta
Trate de utilizar solamente las opciones de respuesta "Sí", "No" y "No se puede valorar". "N/a" se debe utilizar si un estudio no informa una comparación entre los grupos.

13.4.3 Resumen

- La revisión de las citas y los resúmenes identificados mediante la búsqueda consumirá mucho tiempo, en primer lugar debido al volumen de citas identificadas y en segundo lugar debido a que es posible que en el título o en el resumen no se proporcione la información necesaria para evaluar su elegibilidad.
- Obtener los datos de manera similar a un ensayo aleatorizado (es decir, detalles del estudio, la población de estudio, el tamaño de muestra reclutado, el tamaño de muestra analizado, etc.).
- Obtener los datos acerca de lo que hicieron los investigadores (lista de verificación del GMENA o similar).
- Obtener los datos acerca de los factores de confusión considerados.
- Obtener los datos acerca de la comparabilidad de los grupos o los factores de confusión considerados.
- Obtener los datos acerca de los métodos utilizados para controlar por los factores de confusión.
- Obtener los datos acerca de múltiples estimaciones del efecto (estimaciones ajustadas y no ajustadas, si están disponibles).

13.5 Evaluación del riesgo de sesgo en los estudios no aleatorizados

13.5.1 ¿Qué es diferente cuando se incluyen estudios no aleatorizados?

13.5.1.1 Fuentes de sesgo en los estudios no aleatorizados

El sesgo puede estar presente en los hallazgos de los ENA de una manera muy similar a como ocurre en los ensayos aleatorizados diseñados o realizados de manera deficiente (ver Capítulo 8). Por ejemplo, con frecuencia en los ENA no está claro el número de exclusiones, la intervención y la evaluación del resultado no se realizan según protocolos estandarizados y es posible que los resultados no se evalúen de forma cegada. Es probable que los sesgos provocados por estos problemas sean similares a los que ocurren en los ensayos aleatorizados, y los revisores deberían familiarizarse con el Capítulo 8 que describe estos aspectos. Ninguno de estos problemas es menos difícil de superar en un estudio no aleatorizado prospectivo bien planificado que en un ensayo aleatorizado.

En los ENA el uso de mecanismos de asignación diferentes de la asignación al azar oculta da lugar a que sea poco probable que los grupos sean comparables. Estas diferencias sistemáticas potenciales entre las características de los participantes en los diferentes "grupos" de intervención pueden ser un aspecto preocupante clave en la mayoría de los ENA, y aquí se les llama sesgo de selección. Cuando el sesgo de selección produce desequilibrios en factores pronósticos asociados con el resultado de interés, se dice que existen "factores de confusión". En ocasiones se utilizan métodos estadísticos para tener en cuenta el sesgo introducido por los factores de confusión, al producir estimaciones "ajustadas" de los efectos de la intervención y parte de la evaluación de la calidad del estudio incluye considerar lo apropiado del análisis, así como el diseño y la realización del estudio.

La variedad de los diseños de los estudios clasificados como ENA y su diferente susceptibilidad a los diferentes sesgos hace difícil producir una herramienta genérica sólida que se pueda utilizar para evaluar el riesgo de sesgo. En una revisión que incluya ENA con diseños diferentes, puede ser necesario crear varias herramientas para la evaluación del riesgo de sesgo. La inclusión de un metodólogo con experiencia en el equipo de revisión es fundamental para identificar las áreas clave de debilidad en los diseños de los estudios incluidos.

Con los ensayos aleatorizados la evaluación del riesgo de sesgo se centra en el sesgo sistemático, que habitualmente se supone que tiene una dirección "optimista". La tendencia de los investigadores a diseñar, realizar, analizar y describir sus estudios primarios de manera que den resultados esperados, de manera consciente o inconsciente, también es probable que se aplique a los ENA cuando los investigadores tienen el control de decisiones clave (p.ej. asignación a la intervención o selección de los centros). En los ENA verdaderamente observacionales, es posible que el sesgo que surge de los "factores de confusión por indicación" no sea tan consistente; los profesionales de la atención sanitaria pueden tener opiniones diferentes acerca de lo apropiado de las intervenciones alternativas para sus pacientes, que dependen de la gravedad actual de la enfermedad de los pacientes y las comorbilidades. Las diferencias en el case-mix entre los lugares que se comparan pueden ser fortuitas. Por lo tanto, cuando se revisan ENA, la variabilidad de los sesgos y la heterogeneidad entre estudios que los mismos inducen es al menos tan importante como el sesgo sistemático.

13.5.1.2 Evidencia de riesgo de sesgo en los estudios no aleatorizados

Es posible obtener cierta percepción del riesgo de sesgo en los estudios no aleatorizados al comparar ensayos aleatorizados con bajo riesgo de sesgo con ensayos aleatorizados con alto riesgo de sesgo. Los ensayos controlados que asignan participantes mediante procedimientos cuasialeatorizados o que no pueden ocultar la asignación durante el reclutamiento, presentan un riesgo de sesgo de selección similar al de un ensayo o un estudio de cohortes abiertamente no aleatorizado realizado de forma prospectiva. El Capítulo 8 revisa la evidencia sobre varios aspectos del riesgo de sesgo en los ensayos aleatorizados y señala que las limitaciones metodológicas en los ensayos aleatorizados tienden a exagerar los efectos beneficiosos de las intervenciones.

Los investigadores también han comparado los hallazgos de metanálisis separados de ensayos aleatorizados y ENA con la misma pregunta de investigación, bajo la suposición de que estas revisiones sistemáticas metodológicas proporcionan una manera de investigar el riesgo de sesgo en los ENA. Algunas revisiones de este tipo han mostrado discrepancias según el diseño del estudio pero es muy difícil realizar comparaciones justas (McLehose 2000). Hay al menos dos motivos para ello:

- Los ensayos aleatorizados y los ENA con exactamente la misma pregunta son poco frecuentes; por ejemplo, los estudios de la misma intervención que utilizan diferentes

diseños de estudio habitualmente difieren de forma sistemática con respecto a la población, la intervención o el resultado.
- Los ensayos aleatorizados y los ENA pueden diferir sistemáticamente de varias formas con respecto a su riesgo de sesgo (sesgo de notificación, así como de selección, realización, detección y desgaste), y con frecuencia los ENA tiene una calidad relativamente deficiente.

Estos motivos pueden explicar la falta de conclusiones consistentes de las revisiones sistemáticas que han comparado hallazgos de ensayos aleatorizados y ENA con la misma pregunta de investigación. Deeks y cols. revisaron ocho de estas revisiones (Deecks 2003) y encontraron que:

- 5/8 concluyeron que había diferencias entre las estimaciones del efecto de los ensayos aleatorizados y los ENA para muchas pero no todas las intervenciones, sin un patrón consistente;
- 1/8 concluyó que los ENA sobrestimaron el efecto (beneficio) de todas las intervenciones estudiadas;
- 2/8 concluyeron que las estimaciones del efecto en los ensayos aleatorizados y los ENA fueron “marcadamente similares”.

Una revisión metodológicamente similar comparó los hallazgos de ensayos aleatorizados y estudios de preferencia del paciente (King 2005). La revisión concluyó que hay pocas pruebas de que las preferencias “afectaron significativamente la validez”, por lo que dichas preferencias no parecieron ser factores de confusión para los efectos de la intervención.

Hay algunas consideraciones relevantes en la interpretación de este tipo de estudios empíricos. Primero, la publicación de los estudios primarios y su selección por los revisores pueden estar sesgadas. También existe la posibilidad de sesgo en la clasificación de los hallazgos de la revisión. Deeks y cols. encontraron que la misma comparación se clasificó algunas veces como discrepante en una revisión y comparable en una segunda. Lo anterior destaca la dificultad para definir lo que representa una “diferencia”.

En segundo lugar, la observación de que las diferencias no fueron consistentemente optimistas es importante y consistente con el principio de que las estimaciones de los ENA son más heterogéneas que lo esperado por el azar (Greenland 2004). Algunas pruebas empíricas para lo anterior provienen de estudios de simulación innovadora (Deeks 2003). Deeks y cols. señalaron que los sesgos en los ENA son muy variables, y puede ser mejor considerar que introducen una incertidumbre adicional en los resultados, en lugar de un sesgo sistemático estimable. Esta incertidumbre influye adicionalmente a la que se toma en cuenta en los intervalos de confianza y en estudios grandes puede tener fácilmente 5 a 10 veces la magnitud del intervalo de confianza del 95%.

Finalmente, las revisiones metodológicas caen en un círculo cerrado: necesitan suponer que los ENA son válidos y que por lo tanto las diferencias entre las estimaciones del efecto de los ensayos aleatorizados y los ENA también son válidas y se pueden atribuir a factores externos, o que los ENA están segados y por lo tanto las diferencias en las estimaciones del efecto de los ensayos aleatorizados y los ENA se pueden explicar por el riesgo de sesgo diferencial. La verdad puede estar en cualquier lugar dentro de estos dos extremos, pero se mantiene el hecho de que las revisiones metodológicas no pueden de forma inequívoca dividir las discrepancias entre diferentes fuentes. Además, si múltiples factores diferencian a los ensayos aleatorizados de los ENA e influyen en el tamaño del efecto, entonces la no observación de diferencias entre las estimaciones de los tamaños del efecto de los ensayos aleatorizados y los ENA también se puede explicar como consecuencia de los efectos de múltiples factores que influyen en el efecto de una intervención en diferentes direcciones. No es lógico suponer que el hallazgo de una falta de diferencias signifique que los ENA son válidos, ni que el hallazgo de una diferencia signifique que los ENA no son válidos.

13.5.2 Recomendaciones y recursos disponibles para apoyar a los revisores

13.5.2.1 Consideraciones generales al evaluar el riesgo de sesgo en los estudios no aleatorizados

La publicación de los ensayos aleatorizados es relativamente sencilla y se guía cada vez más por la declaración CONSORT (Moher 2001). Mucho más recientemente se ha desarrollado una declaración de consenso similar, STROBE, para la publicación de estudios epidemiológicos observacionales (Vandenbroucke 2007, von Elm 2007). Por lo tanto, es probable que la calidad de la información necesaria para evaluar el riesgo de sesgo no sea tan buena para los ENA. Es posible que lo anterior dificulte cualquier evaluación del riesgo de sesgo.

Un protocolo es una herramienta que protege del sesgo; cuando se registra antes de que el estudio comience, prueba que los aspectos del diseño y el análisis del estudio se consideraron antes de comenzar el reclutamiento y que se especificaron las definiciones de los datos y los métodos para estandarizar su obtención. Debido a la necesidad de la aprobación ética de la investigación, todos los ensayos aleatorizados deben tener un protocolo, incluso cuando los protocolos varían en su calidad y los ítems que especifican. Muchos ensayos aleatorizados, particularmente los patrocinados por la industria, también tienen manuales de estudio detallados. Históricamente los investigadores no han tenido que obtener aprobación ética de investigación para los ENA, y los ENA primarios pocas veces notifican si los métodos se basan en un protocolo. Por lo tanto, la protección ofrecida por un protocolo a menudo no existe para los ENA. Las implicaciones de no tener un protocolo no se han investigado. Sin embargo, lo anterior significa, por ejemplo, que no hay limitaciones para la tendencia de los investigadores de “cosechar” resultados, subgrupos y análisis, lo que sucede en mayor o menor grado incluso en los ensayos aleatorizados en los que existen protocolos (Chan 2004).

Al igual que en los ensayos aleatorizados, las dimensiones del sesgo a evaluar incluyen el sesgo de selección (relacionado con la comparabilidad de los grupos, los factores de confusión y el ajuste); el sesgo de realización (relacionado con la fidelidad de las intervenciones y la calidad de la información con respecto a quién recibió qué intervenciones, incluido el cegamiento de los participantes y y del personal sanitario), el sesgo de detección (relacionado con la evaluación no sesgada y adecuada del resultado, incluido el cegamiento de los evaluadores), el sesgo de desgaste (relacionado con la compleción de la muestra, el seguimiento y los datos) y el sesgo de notificación (relacionado con los sesgos de publicación y la notificación selectiva de los resultados). La evaluación del riesgo de sesgo en los ensayos aleatorizados se ha desarrollado identificando las características del diseño que se utilizan para evitar cada una de estas dimensiones, y señalar si cumplen los requisitos. Las evaluaciones del riesgo de sesgo para los ENA deberían proceder de la misma manera, con la preespecificación de las características a evaluar en el protocolo y registrando lo que sucedió en el estudio, así como una valoración de si el mismo fue adecuado, inadecuado o incierto como una forma de evitar el riesgo de este sesgo particular. Es probable que para determinar estas características se requiera el aporte experto de un epidemiólogo y también dependerá en parte de la pregunta clínica. Se debería prestar especial atención a la evaluación de los factores de confusión (ver Sección 13.5.2.2).

El motivo para prestar una cuidadosa atención a las *características* del diseño de los estudios primarios (por ejemplo, cómo se asignaron los participante a los grupos, o qué partes del estudio fueron prospectivas) en lugar de a las *calificaciones* del diseño (como “cohortes” o “transversal”) se debe a que se ha establecido la hipótesis de que el riesgo de sesgo está influenciado por las características específicas de un estudio, en lugar de por una categorización amplia del enfoque utilizado. Además, los términos como “cohortes” y “transversal” son ambiguos y cubren un amplio rango de diseños específicos de estudios.

No hay una lista derivada de estudios empíricos disponible para las características del diseño del estudio que sea relevante para el riesgo de sesgo, aunque es posible construir una lista corta a partir de la evidencia y la teoría acerca del riesgo de sesgo en los estudios etiológicos y los ensayos aleatorizados (ver Secciones 13.2.2 y 13.4.2).

Debido a la diversidad de ENA, pueden ser necesarios varios métodos para evaluar los ENA con diferentes características de diseño. Una diferencia importante se encuentra entre los estudios en los cuales la asignación a los grupos se realiza por el resultado (p.ej. estudios de casos y controles) y los estudios cuya asignación a los grupos se relaciona de forma más directa con las intervenciones. En el primer tipo de estudios, es la exposición de interés más que el resultado la que es más susceptible al sesgo; los revisores deberían preguntarse si los investigadores que evaluaron la exposición estaban cegados a si los participantes habían presentado el resultado o no (es decir, si eran casos o controles). Los estudios de casos y controles son muy adecuados para investigar asociaciones entre resultados poco frecuentes y múltiples exposiciones, por lo que tienen una función importante en la generación de pruebas acerca de efectos adversos potenciales y efectos beneficiosos inesperados de las intervenciones. También se han utilizado para evaluar intervenciones de salud pública a gran escala como la prevención de accidentes y el cribado (MacLehose 2000), que son difíciles o costosas de evaluar mediante ensayos aleatorizados. Sin embargo, los revisores se deberían familiarizar con las consideraciones epidemiológicas que se aplican particularmente a dichos estudios (Rothman 1986). Es de señalar que algunos análisis de registros de pacientes también tienen similitudes con los estudios de casos y controles; por ejemplo, si la base de datos completa se divide en grupos de pacientes que han presentado o no un determinado resultado y se investigan las exposiciones asociadas con el resultado. Los revisores necesitan un conocimiento más profundo de epidemiología cuando evalúan el riesgo de sesgo en los ENA, comparados con los ensayos aleatorizados.

13.5.2.2 Factores de confusión y ajuste

Los investigadores no siempre toman las mismas decisiones con respecto a los factores de confusión, por lo que el método utilizado para controlar por los factores de confusión es una importante fuente de heterogeneidad entre los estudios. Puede haber diferencias en los factores de confusión considerados, el método utilizado para controlar por los factores de confusión y la forma precisa mediante la cual es posible medir los factores de confusión e incluirlos en los análisis. Muchos (pero no todos) los ENA describen los factores de confusión considerados y si los mismos se tomaron en cuenta en el diseño o el análisis del estudio; la mayoría también describe las características iniciales de los grupos que se comparan. Sin embargo, puede ser difícil evaluar qué hicieron realmente los investigadores para controlar por los factores de confusión; muy pocos estudios describen con exactitud cómo se midieron los factores de confusión o cómo se ajustaron como covariables en los modelos de regresión (p.ej. como una variable continua, ordinal o categórica agrupada).

Algunas indicaciones específicas para evaluar el riesgo de sesgo de selección son las siguientes:

- En el estadio de redacción del protocolo, enumerar los posibles factores de confusión.
- Identificar los factores de confusión que los investigadores han considerado y los que se han omitido. Señalar cómo se han medido (la capacidad de controlar por un factor de confusión depende de la precisión con la cual se midió el factor).
- Evaluar el equilibrio entre los grupos comparadores al inicio con respecto a los factores pronósticos o de confusión principales.
- Identificar lo que hicieron los investigadores para controlar el sesgo de selección, es decir, cualquier característica del diseño utilizada para este fin (p.ej. pareamiento o limitación a subgrupos determinados) y los métodos de análisis (p.ej. estratificación o modelo de regresión con puntuaciones de propensión o covariables).

No hay un método establecido para identificar un grupo preespecificado de factores de confusión importantes. La enumeración de posibles factores de confusión se debe hacer "de forma independiente" y, se pudiera argumentar, "sistemáticamente". La lista no se debe generar solamente sobre la base de los factores considerados en los estudios primarios incluidos en la revisión (al menos, no sin algún formulario de validación independiente), ya que el número de posibles factores de confusión es probable que aumente con el tiempo (por lo tanto, los estudios más antiguos pueden no estar actualizados) y los propios investigadores pueden sencillamente decidir medir factores de confusión considerados en estudios previos (por lo tanto, la lista puede ser selectiva). Los investigadores que estudian asociaciones etiológicas a menudo no explican su selección de factores de confusión (Pocock 2004). En su lugar, la lista se debe basar en la evidencia (aunque la realización de una revisión sistemática para identificar todos los posibles factores pronósticos es un recurso extremo) y en la opinión experta de los miembros del equipo de revisión y los asesores.

La descripción de las evaluaciones de los factores de confusión en una revisión Cochrane se puede lograr mejor al crear tablas adicionales que enumeren los factores de confusión preestablecidos en las columnas, con los estudios en las filas y que indiquen si cada estudio (i) limitó la selección de participantes de manera que todos los grupos tenían el mismo valor para el factor de confusión (p.ej. al limitar el estudio a los participantes masculinos solamente), (ii) mostró un equilibrio entre los grupos en cuanto a los factores de confusión, (iii) pareó por factor de confusión o (iv) ajustó por el factor de confusión en los análisis estadísticos para cuantificar el tamaño del efecto.

13.5.2.3 Herramientas para evaluar la calidad metodológica o el riesgo de sesgo en los estudios no aleatorizados

El Capítulo 8 (Sección 8.5) describe la herramienta "Riesgo de sesgo" que se espera que los revisores utilicen par evaluar el riesgo de sesgo en los ensayos aleatorizados. Este incluye la consideración de seis características: generación de la secuencia, ocultación de la secuencia de asignación, cegamiento, datos de resultado incompletos, notificación selectiva de resultado y "otras" fuentes potenciales de sesgo. Los ítems se evalúan de acuerdo a: (i) si proporcionan una descripción de lo que sucedió en el estudio, (ii) si proporcionan una valoración de lo adecuado del estudio con respecto al ítem. La valoración se realiza al responder una pregunta preespecificada, por lo que una respuesta "Sí" indica bajo riesgo de sesgo, una respuesta "No" indica alto riesgo de sesgo y una respuesta "Incierto" indica que el riesgo de sesgo es incierto o desconocido. La herramienta no se desarrolló teniendo en cuenta los ENA, y los seis dominios no son necesariamente apropiados para los mismos. Sin embargo, al podría ser útil seguir la estructura general de la herramienta y las evaluaciones cuando se crean instrumentos de evaluación del riesgo de sesgo para los ENA.

Para los estudios experimentales y controlados y para los estudios de cohortes prospectivas (ver Recuadro 13.1.a y Sección 13.2.2) podría ser útil evaluar los seis dominios de la herramienta "Riesgo de sesgo", sea la asignación al azar o no. Ésta es la evaluación mínima que los revisores deberían hacer y habitualmente se necesitan más detalles. Un componente adicional es evaluar el riesgo de sesgo debido a los factores de confusión. Es probable que la profundidad de esta evaluación dependa de la heterogeneidad entre los estudios y de si los revisores proponen una síntesis cuantitativa (ver Sección 13.6). Si los estudios son heterogéneos y no se propone una síntesis cuantitativa, entonces una evaluación menos detallada puede servir para ilustrar la heterogeneidad y detallar la interpretación de los hallazgos de la revisión.

Se han creado muchas herramientas para evaluar la calidad metodológica de los estudios no aleatorizados de intervenciones, que han sido revisadas sistemáticamente por Deeks y cols. (Deeks 2003). En su revisión localizaron 182 herramientas, que ellos redujeron a una lista corta de 14, e identificaron seis como potencialmente útiles para las revisiones

sistemáticas, ya que "obligan a los revisores a ser sistemáticos en sus evaluaciones de los estudios e intentar asegurar que se hagan valoraciones de calidad de la forma más objetiva posible". Sin embargo, las seis necesitaron ciertos ajustes porque descuidaron la obtención de información detallada acerca de cómo los participantes de los estudios se asignaron a los grupos, lo que es probable que sea fundamental en términos del riesgo de sesgo de selección. De estas seis herramientas, no todas fueron adecuadas para los diferentes diseños de estudios. En común con otras herramientas para evaluar la calidad de los ensayos aleatorizados, algunas no diferencian los ítems relacionados con la calidad del estudio y la calidad de la descripción del estudio. Las dos herramientas más útiles identificadas en esta revisión son el instrumento de Downs y Black y la Escala Newcastle-Ottawa (Downs 1998, Wells 2008).

El instrumento de Downs y Black se ha modificado para su uso en revisiones sistemáticas metodológicas (MacLehore 2000). Los revisores encontraron que algunos de los 29 ítems fueron difíciles de aplicar a los estudios de casos y controles, que el instrumento necesita una experiencia epidemiológica considerable y que su uso consume tiempo. La Escala Newcastle-Ottawa, que se ha utilizado en talleres del GMENA para ilustrar aspectos de la extracción de datos de ENA primarios, contiene solamente ocho ítems y es más sencilla de aplicar (Wells 2008). Sin embargo, es posible que los ítems aún necesiten adaptarse a la pregunta de investigación de interés. Los revisores también deberían conocer las diferencias en la terminología epidemiológica en diferentes países; por ejemplo, la Escala Newcastle-Ottawa utiliza el término "sesgo de selección" para describir lo que otros pueden llamar "aplicabilidad" o "generalizabilidad".

Al reconocer la importancia de distinguir entre "lo que los investigadores hacen" y "lo que los investigadores informan" los revisores también pueden encontrar útil considerar ítems incluidos en normas de publicación para ensayos aleatorizados (Moher 2001) y estudios epidemiológicos observacionales (Vandenbroucke 2007) con el fin de destacar las brechas en la publicación (y la realización) de los ENA (Reeves 2004, Reeves 2007).

13.5.2.4 Limitaciones prácticas al evaluar el riesgo de sesgo en los estudios no aleatorizados

Dos estudios de revisiones sistemáticas que incluyeron ENA han comentado que sólo una minoría de revisiones evaluaron la calidad metodológica de los estudios incluidos (Audige 2004, Golder 2006a). Los miembros del GMENA han ganado experiencia en tratar de evaluar el riesgo de sesgo de los estudios no aleatorizados. De manera anecdótica, los autores de estas revisiones señalan que generalmente los ENA tienen una calidad metodológica deficiente o su redacción también es deficiente, por lo que la evaluación de la calidad metodológica y el riesgo de sesgo de manera consistente entre los estudios primarios es difícil o imposible (Kwuan 2004). Incluso se ha notificado que la escala Newcastle-Ottawa es difícil de aplicar, y es probable que el acuerdo entre revisores sea modesto. La información metodológica puede ser difícil de encontrar en los artículos, lo que hace que la tarea sea frustrante, especialmente cuando se utiliza alguno de los instrumentos más detallados; los revisores pueden pasar un largo tiempo en la búsqueda de detalles de lo que hicieron los investigadores, sólo para concluir que la información no se proporcionó. Sin embargo, la obtención de alguna información factual (por ejemplo, los factores de confusión considerados y lo que hicieron los investigadores con respecto a dichos factores) aún puede ser útil, ya que dicha información ilustra el grado de heterogeneidad entre los estudios.

13.5.3 Resumen

- En el estadio de redacción del protocolo para la revisión, compilar una lista de posibles factores de confusión y justificar la selección.
- En el estadio de redacción del protocolo para la revisión, decidir cómo se evaluará el riesgo de sesgo en los estudios primarios, incluido el grado de control de los factores de confusión.
- Para los ENA realizados completamente de forma prospectiva, aplicar los métodos que la Colaboración recomienda para los ensayos aleatorizados.
- No se recomienda un instrumento único, por lo que es probable que los revisores necesiten incluir instrumentos o ítems complementarios para la evaluación del riesgo de sesgo.
- Los aspectos como los factores de confusión no se pueden abordar fácilmente dentro del formato de la nueva herramienta de riesgo de sesgo y requieren la creación de tablas adicionales para detallar las evaluaciones.
- La obtención de alguna información factual (por ejemplo, los factores de confusión considerados y lo que hicieron los investigadores con respecto a dichos factores) es útil, ya que dicha información ilustra el grado de heterogeneidad entre los estudios
- Los revisores que deciden incluir estudios de casos y controles en una revisión Cochrane deberían asegurarse de estar familiarizados con las deficiencias frecuentes que pueden afectar a dichos estudios y evaluar su susceptibilidad al sesgo mediante un instrumento diseñado para ese fin.
- Los revisores pueden decidir que no se justifica obtener muchos detalles acerca del riesgo de los factores de confusión y de otros sesgos. Sin embargo, si se adopta este enfoque, los revisores deberían reconocer el grado potencial de heterogeneidad entre los estudios con respecto a los posibles factores de confusión residuales y otros sesgos, y demostrar que consideraron esta fuente de heterogeneidad cuando interpretaron los hallazgos de los ENA primarios revisados.

13.6 Síntesis de los datos de los estudios no aleatorizados

13.6.1 ¿Qué es diferente cuando se incluyen estudios no aleatorizados?

Los revisores deberían esperar una mayor heterogeneidad en una revisión sistemática de ENA que en una revisión sistemática de ensayos aleatorizados. Esto se debe al aumento potencial de diversidad metodológica debido a la variación entre los estudios primarios en cuanto a su riesgo de sesgo de selección, la variación en la forma en la cual se consideran los factores de confusión en el análisis y un mayor riesgo de otros sesgos debido al diseño y la realización deficientes. No hay forma de controlar por estos sesgos en el análisis de los estudios primarios ni un método establecido para evaluar cómo, o el grado en el cual, estos sesgos afectan los estudios primarios (no obstante, ver Capítulo 8).

Existe un grupo de opinión que cree apropiado combinar los resultados de los estudios no aleatorizados cuando los mismos tienen efectos grandes, pero la lógica de este punto de vista puede ser cuestionada. Es tan probable (o más probable) que los ENA con efectos grandes estén sesgados y sean heterogéneos, comparados con los ENA con efectos pequeños. Las valoraciones acerca del riesgo de sesgo y la heterogeneidad se deberían basar en una evaluación crítica de las características y los métodos de los estudios incluidos, no en sus resultados.

Cuando se evalúa la similitud de los estudios antes del metanálisis, los revisores también deberían recordar que algunas características de los estudios, como la evaluación del

resultado no enmascarada a la asignación a la intervención, pueden ser relativamente homogéneas entre los ENA pero aún se mantiene el riesgo de sesgo de todos los estudios.

Si los revisores consideran que los ENA incluidos son razonablemente resistentes a los sesgos y relativamente homogéneos en este aspecto, pueden desear combinar los datos de los estudios mediante un metanálisis (Taggart 2001). Al contrario que para los ensayos aleatorizados, habitualmente será apropiado analizar las estimaciones del efecto ajustadas, en lugar de no ajustadas; es decir, realizar análisis que intentan "controlar por los factores de confusión". Esto puede requerir que los revisores escojan entre estimaciones ajustadas alternativas descritas en un estudio. El metanálisis de las estimaciones ajustadas se puede realizar como un promedio ponderado de la varianza inversa, por ejemplo, mediante el tipo de resultado "varianza inversa genérica" de RevMan (ver Capítulo 9, Sección 9.4.3). En principio, cualquier medida del efecto utilizada en el metanálisis de ensayos aleatorizados se puede utilizar en el metanálisis de estudios no aleatorizados (ver Capítulo 9, Sección 9.2), aunque habitualmente se utilizará el *odds ratio* porque es la única medida del efecto para resultados dicotómicos que se puede calcular en los estudios de casos y controles, y se calcula cuando se utiliza la regresión logística para ajustar por los factores de confusión.

Un peligro es que los ENA muy grandes de calidad metodológica deficiente (por ejemplo, basados en datos que se obtienen habitualmente) pueden dominar los hallazgos de otros estudios pequeños con menos riesgo de sesgo (en los que posiblemente la obtención de los datos fue adaptada). Los revisores deberían recordar que los intervalos de confianza para las estimaciones del efecto de ENA grandes es menos probable que representen la verdadera incertidumbre del efecto observado, en comparación con los intervalos de confianza de los ENA más pequeños (ver Sección 13.5.1.2), aunque no hay forma de calcular o corregir este efecto.

13.6.2 Recomendaciones y recursos disponibles para apoyar a los revisores

13.6.2.1 Control por factores de confusión

Los desequilibrios en los factores pronósticos en los ENA (p.ej. "factores de confusión por indicación" (Grobbee 1997) se deben tomar en cuenta en el análisis estadístico. Hay varios métodos para controlar por los factores de confusión. El pareamiento, es decir, la generación de grupos de intervención similares con respecto a factores pronósticos importantes, se pueden utilizar para disminuir los factores de confusión en el estadio de diseño del estudio. La estratificación y el modelo de regresión son enfoques estadísticos para controlar por los factores de confusión, lo que da lugar a una estimación del efecto de la intervención ajustada para los desequilibrios en los factores pronósticos observados. Algunos análisis utilizan métodos de puntuación de propensión como parte de un análisis de dos etapas. La probabilidad de que un individuo reciba la intervención experimental (la puntuación de propensión) se calcula primero según sus características mediante un modelo de regresión logística. Esta medida resumen sencilla del case-mix se utiliza luego para el pareamiento, la estratificación o en un modelo de regresión.

Pareamiento

La selección de pacientes con valores similares para los factores pronósticos importantes da lugar a grupos más comparables. Por lo tanto, el pareamiento se puede considerar un tipo de ajuste por factores de confusión. El pareamiento se puede realizar a nivel de los pacientes individuales (es decir, se seleccionan uno o más participantes control que tiene características similares a las de los participantes de una intervención) o nivel del estrato de los participantes (es decir, los participantes se seleccionan de manera que haya el mismo número de participantes control en un estrato, por ejemplo, 60 años o más, similar al grupo de intervención). Cuando se utiliza el pareamiento directo se debe considerar en el análisis estadístico de un único estudio la naturaleza pareada de los datos, con el fin de

obtener intervalos de confianza apropiados para la estimación del efecto de la intervención. El pareamiento con respecto a una medida única como la puntuación de propensión es más fácil de lograr que el pareamiento de los individuos con un grupo determinado de características.

Estratificación

La estratificación implica la división de los participantes en subgrupos con respecto a factores pronósticos categóricos (o categorizados como cuantitativos), por ejemplo clasificar la edad en décadas, o el peso en cuartiles. El efecto de la intervención se calcula luego en cada estrato y se calcula una estimación combinada entre los estratos. Este procedimiento se puede interpretar como un metanálisis a nivel de un estudio individual. Para los resultados dicotómicos se utiliza a menudo el método de Mantel-Haenszel para calcular el efecto general de la intervención, con versiones disponibles para el *odds ratio*, el riesgo relativo y la diferencia de riesgos como medidas del efecto de la intervención. Nuevamente, la puntuación de propensión se pude utilizar como la variable de estratificación.

Modelado

En un enfoque de modelado, la información sobre la intervención y los factores pronósticos se incorpora a una ecuación de regresión. Las ventajas de los modelos de regresión incluyen la posibilidad de incorporar factores cuantitativos sin categorización y la posibilidad de modelar tendencias en los factores de confusión medidos en una escala ordinal. Para los resultados dicotómicos casi siempre se utiliza un modelo de regresión logística para calcular el efecto ajustado de la intervención. Por lo tanto, se utiliza (implícitamente) el *odds ratio* como la medida del efecto de la intervención. Los modelos de regresión también están disponibles para el riesgo relativo y la reducción del riesgo absoluto como medidas del efecto, pero pocas veces estos modelos se utilizan en la práctica. Habitualmente se utiliza un modelo de regresión lineal para los resultados continuos (probablemente después de la transformación de una o más variables) y habitualmente se utiliza un modelo de regresión de riesgos proporcionales (regresión de Cox) para los datos de tiempo hasta el evento. Los modelos de regresión también pueden utilizar la puntuación de propensión sola o en combinación con otras características de los participantes como variables explicativas.

Los revisores deberían reconocer que en cualquier estudio no aleatorizado, incluso cuando los grupos experimental y control parecen comparables al inicio, la estimación del tamaño del efecto aún tiene riesgo de sesgo debido a los factores de confusión residuales. Esto se debe a que todos los métodos para controlar por los factores de confusión son imperfectos, por ejemplo, por los siguientes motivos:

- Factores de confusión desconocidos y por lo tanto no medidos, que no se pueden controlar.
- Resolución deficiente en la medición de los factores de confusión, p.ej. morbilidad evaluada en una escala ordinal sencilla (Concato 1992), que representa una mala clasificación no diferencial del error con respecto a los factores de confusión.
- Limitaciones prácticas en la resolución del pareamiento, y el número de factores de confusión en los cuales los participantes se pueden parear, en los análisis pareados.
- Resolución deficiente en la forma en la que se miden los factores de confusión en los análisis estratificados, o se tratan en los análisis, ilustrado por la amplitud del estrato (p.ej. décadas de edad); esta limitación también se aplica a los modelos de regresión cuando los factores de confusión se categorizan y modelan de forma discreta.
- Suposiciones en la forma en la que los factores de confusión se modelan en los análisis de regresión debido al conocimiento imperfecto de la forma de asociación entre el factor de confusión y el resultado.
- No hay un método establecido para valorar el alcance probable de los factores de confusión residuales. La dirección del sesgo de los factores de confusión es impredecible y puede diferir entre los estudios.

13.6.2.2 Combinación de los estudios

Es de esperar que los efectos de la intervención calculados para diferentes diseños de estudios estén influenciados en mayor o menor grado por las diferentes fuentes de sesgo (ver Sección 13.5). Se debe esperar que los resultados de los diferentes diseños de estudios difieran sistemáticamente, lo que dará lugar a un aumento de la heterogeneidad. Por lo tanto, se recomienda que los ENA que utilicen diferentes diseños de estudios (o que tengan diferentes características en el diseño), o los ensayos aleatorizados y los ENA, no se combinen en un metanálisis.

Debido a la necesidad de controlar lo mejor posible por los factores de confusión, la estimación del efecto de la intervención y su error estándar (o intervalo de confianza) son piezas de información clave que se deberían utilizar para combinar los ENA en un metanálisis. Los numeradores y denominadores únicos, o las medias y los errores estándar, para los grupos intervención y control no se pueden controlar por los factores de confusión a menos que los grupos se hayan pareado en el estadio de diseño. Por lo tanto, los métodos de metanálisis basados en las estimaciones y los errores estándar, y en particular el método de la varianza inversa genérica, serán adecuados para los ENA (ver Capítulo 9, Sección 9.4.3).

Es sencillo extraer una estimación del efecto ajustada y su error estándar para un metanálisis si se describe una estimación ajustada única para un determinado resultado en un ENA primario. Sin embargo, muchos ENA describen estimaciones del efecto ajustadas y no ajustadas, y algunos ENA notifican estimaciones múltiples ajustadas de los análisis, incluidos diferentes grupos de covariables. Los revisores deberían registrar las estimaciones ajustadas y no ajustadas de los efectos, pero puede ser difícil elegir entre estimaciones ajustadas alternativas. No es posible hacer recomendaciones generales para la selección de una estimación ajustada preferible. Las posibles reglas de selección son:

- utilizar la estimación del modelo que ajustó por el número máximo de covariables;
- utilizar la estimación que los autores identifican como el modelo primario ajustado; y
- utilizar la estimación del modelo que incluya el mayor número de factores de confusión considerados importantes al inicio por los revisores.

Se pueden realizar análisis de sensibilidad al combinar por separado los resultados más optimistas y pesimistas de cada estudio incluido.

Hay un aspecto estadístico sutil con respecto a la interpretación diferente de los efectos ajustados y no ajustados cuando se expresan como *odds ratios* o cocientes de riesgos instantáneos (CRI). La estimación del efecto no ajustada se conoce como el efecto promedio en la población, y si la estimación no estuviera sesgada sería el efecto de la intervención observado en una población con una mezcla promedio de características de pronóstico. Cuando las estimaciones se ajustan por las características de pronóstico, las estimaciones del efecto se conocen como estimaciones condicionales y son los efectos de la intervención que se observarían en grupos con combinaciones particulares de las covariables ajustadas. La investigación matemática ha mostrado que habitualmente las estimaciones condicionales son más grandes (más allá de un OR o un CRI de 1), que las estimaciones promedio de la población. Es posible que este fenómeno no se observe en las revisiones sistemáticas debido a la heterogeneidad en las estimaciones de los estudios.

13.6.2.3 Análisis de la heterogeneidad

La exploración de posibles fuentes de heterogeneidad entre los estudios debe ser parte de cualquier revisión Cochrane y se discute en detalle en el Capítulo 9 (Sección 9.6). Es de esperar que los estudios no aleatorizados sean más heterogéneos que los ensayos aleatorizados, debido a las fuentes adicionales de diversidad metodológica y sesgo. La

manera más simple de mostrar la variación en los resultados de los estudios es mediante un gráfico de bosque (ver Capítulo 11, Sección 11.3.2).

Puede ser de valor realizar análisis de metarregresión para identificar determinantes importantes de heterogeneidad, incluso en revisiones en las que los estudios se consideren demasiado heterogéneos para combinarlos. Dichos análisis pueden ayudar a identificar características metodológicas que se relacionan sistemáticamente con los efectos observados de la intervención y ayudan a identificar subgrupos de estudios que es más probable que produzcan estimaciones válidas de los efectos de la intervención.

13.6.2.4 Cuando se considera que la combinación no es apropiada

Antes de realizar un metanálisis, los revisores deberían hacerse a sí mismos la pregunta estándar acerca de si los estudios primarios son "suficientemente similares" para justificar la combinación (ver Capítulo 9). Los gráficos de bosque en RevMan permiten la presentación de las estimaciones y los errores estándar para cada estudio, mediante el tipo de resultado "Varianza inversa genérica". Los metanálisis se pueden eliminar o incluir solamente para subgrupos dentro de un gráfico. Siempre que las estimaciones del efecto de los estudios incluidos se puedan expresar mediante medidas del efecto consistentes, se recomienda que los revisores muestren los resultados de los estudios individuales para los ENA con características similares del diseño del estudio mediante los gráficos de bosque, como una característica estándar. Si no están disponibles medidas del efecto consistentes, entonces se deberían utilizar tablas adicionales para presentar los resultados en un formato sistemático.

Si los estudios incluidos no son suficientemente homogéneos para combinarlos en un metanálisis (lo cual se espera que sea la norma para las revisiones que incluyen ENA), el GMENA recomienda mostrar los resultados de los estudios incluidos en un gráfico de bosque, pero eliminar la estimación combinada. Los estudios se pueden ordenar en el gráfico de bosque (o mostrar en gráficos de bosque separados) según la característica del diseño del estudio o alguna otra característica que se considere que refleja la susceptibilidad al sesgo (p.ej. el número de "estrellas" en la Escala Newcastle-Ottawa (Wells 2008)). Los diagnósticos e investigaciones de la heterogeneidad (p.ej. una prueba para la heterogeneidad, la estadística I^2 y los análisis de metarregresión) son de valor, incluso si se hizo una valoración sin calcular una estimación combinada del efecto (Higgins 2003, Siegfried 2003).

Sin embargo, los análisis narrativos son problemáticos porque es difícil establecer o describir los resultados sin que sean selectivos o hagan énfasis en algunos hallazgos y no en otros. De manera ideal, los revisores deberían establecer en el protocolo de la revisión cómo ellos planifican utilizar la síntesis narrativa para informar los hallazgos de los estudios primarios.

13.6.3 Resumen

- La heterogeneidad será mayor en una revisión sistemática de ENA que en una revisión sistemática de ensayos aleatorizados. Por lo tanto, los revisores deberían considerar con mucho cuidado el grado probable de heterogeneidad entre los estudios incluidos cuando decidan combinar los hallazgos de forma cuantitativa (es decir, mediante el metanálisis). Se espera que la combinación de las estimaciones del efecto de los ENA sea la excepción en lugar de la regla.
- Las estimaciones del efecto de los ENA no se deberían combinar con las estimaciones del efecto de los ensayos aleatorizados, o entre ENA que tengan características diferentes de diseño del estudio.
- Los gráficos de bosque se deberían utilizar para resumir los hallazgos de los estudios incluidos.

- El diagnóstico y las investigaciones de la heterogeneidad se pueden utilizar independientemente de si se ha tomado o no la decisión de combinar las estimaciones del efecto de estudios diferentes.

13.7 Interpretación y discusión

13.7.1 Retos en la interpretación de las revisiones Cochrane de efectividad que incluyen estudios no aleatorizados

Introducción

Los revisores se enfrentan a grandes retos para demostrar de manera convincente que el resultado de una revisión Cochrane de ENA puede proporcionar una respuesta cercana a la definitiva acerca del probable efecto de una intervención (Deeks 2003). En muchas situaciones es probable que las revisiones de ENA concluyan que el cálculo de un efecto "promedio" no es útil (Siegfried 2003), que la evidencia de los ENA no son adecuadas para probar la efectividad o el daño (Kwan 2004) y que se deberían realizar ensayos aleatorizados (Taggart 2001).

Los retos surgen en todos los estadios de la realización de una revisión de ENA: en la decisión de qué diseños de estudios incluir, la búsqueda de los estudios, la evaluación de sus sesgos potenciales y en la decisión de combinar los resultados. El revisor debe convencer al lector de la revisión de que estos retos se han abordado de manera adecuada o debe discutir cómo y por qué no se pudieron hacer frente. En esta sección los retos se ilustran con referencia a aspectos que surgieron en las diferentes secciones de este capítulo. La sección "Discusión" de la revisión debe abordar el grado en el cual los retos se vencieron.

13.7.1.1 ¿Se han incluido todos los estudios importantes y relevantes?

Incluso si es posible justificar la selección de los diseños de los estudios elegibles, puede ser difícil mostrar que se han identificado todos los estudios relevantes debido a la indexación deficiente y el uso no consistente de las calificaciones de los diseños de los estudio por parte de los investigadores. Es probable que las estrategias de búsqueda exhaustivas que se centran solamente en la afección de salud y la intervención de interés den lugar a listas muy largas de citas que incluyen relativamente pocos estudios elegibles; por el contrario, las estrategias restrictivas perderán inevitablemente algunos estudios elegibles. En la práctica, los recursos disponibles pueden hacer imposible procesar los resultados de una búsqueda exhaustiva, especialmente porque a menudo los revisores tendrán que leer artículos completos en lugar de resúmenes para determinar la elegibilidad. No se conocen las implicaciones del uso de una estrategia de búsqueda más o menos exhaustiva.

13.7.1.2 ¿Se ha evaluado adecuadamente el riesgo de sesgo de los estudios incluidos?

La interpretación de los resultados de una revisión de ENA debería incluir las consideraciones de la probable dirección y magnitud del sesgo. Los sesgos que pueden afectar los ensayos aleatorizados también pueden afectar los ENA, pero generalmente en un mayor grado. Por ejemplo, con frecuencia el desgaste en los ENA es mayor (y se informa de manera deficiente), pocas veces la evaluación de la intervención y del resultado se realiza según protocolos estandarizados y en pocas ocasiones los resultados están cegados. Demasiado a menudo estas limitaciones de los ENA se ven como una parte de la realización de un ENA y sus implicaciones para el riesgo de sesgo no se consideran de una manera adecuada. Por ejemplo, algunos usuarios de la evidencia pueden considerar que

los ENA que investigan resultados a largo plazo tiene "mejor calidad" que los ensayos aleatorizados de resultados a corto plazo sencillamente en base a su relevancia, sin evaluar su riesgo de sesgo (ver Sección 13.2.1.4).

La evaluación de la magnitud de los factores de confusión en los ENA es especialmente problemática. Los revisores no deberían solamente tener métodos adecuados para la evaluación, sino también obtener e informar adecuadamente los detalles acerca de los factores de confusión considerados por los investigadores y los métodos utilizados para controlar por los factores de confusión. Es posible que la información no esté disponible en las publicaciones de los estudios primarios, lo que impide que los revisores investiguen las diferencias en los métodos de los estudios elegibles y otras fuentes de heterogeneidad que se consideraron probablemente importantes cuando se redactó el protocolo.

Los revisores deberían recordar los siguientes puntos acerca de los factores de confusión:

- La dirección del sesgo introducido por los factores de confusión es impredecible.
- Es probable que los métodos utilizados por los investigadores para controlar por los factores de confusión varíen entre los estudios.
- Se desconoce el grado de los factores de confusión residuales en cualquier estudio particular, y es probable que el mismo varíe entre los estudios.
- Los factores de confusión residuales (y otros sesgos) significan que los intervalos de confianza subestiman la verdadera incertidumbre alrededor de la estimación del efecto.
- Es importante identificar los probables factores de confusión por los cuales no se ha ajustado, así como los factores por los cuales se ha ajustado.

- Los retos descritos anteriormente afectan todas las revisiones sistemáticas de ENA. Sin embargo, los retos pueden ser menos extremos en algunas áreas de la atención sanitaria (p.ej. los factores de confusión pueden ser un problema menor en los estudios observacionales de efectos adversos o a largo plazo, o en algunas intervenciones de prevención de la salud pública primaria).

Una pista para la presencia de sesgo es la marcada heterogeneidad entre los estudios. Aunque la heterogeneidad puede surgir debido a diferencias en los participantes, las intervenciones y las evaluaciones de resultado, se debe considerar seriamente la posibilidad de que el sesgo sea la causa de la heterogeneidad en las revisiones de ENA. Sin embargo, la falta de heterogeneidad no indica ausencia de sesgo, ya que es posible que exista un sesgo consistente en todos los estudios.

¿Es posible predecir la magnitud y la dirección del sesgo? Éste es un tema de investigación actual que intenta recopilar las pruebas empíricas de los factores (como el diseño del estudio y el tipo de intervención) que determinan el tamaño y la dirección de estos sesgos. La capacidad de predecir la magnitud y la probable dirección del sesgo mejoraría mucho la utilidad de la evidencia de las revisiones sistemáticas de ENA. Actualmente hay algunas pruebas de que en algunas circunstancias limitadas es posible predecir al menos la dirección (Henry 2001).

13.7.2 Evaluación de la fuerza de la evidencia proporcionada por revisiones que incluyen estudios no aleatorizados

La "exposición" de la evidencia proveniente de ENA sobre una pregunta de salud particular permite el debate informado acerca de su significado e importancia, y de la certidumbre que se les puede atribuir. Críticamente, es necesario que haya un debate acerca de la posibilidad de que los hallazgos observados sean erróneos. Todas las jerarquías de pruebas colocan a los ENA en los últimos lugares de la lista, pero por encima de la opinión clínica (Eccles 1996, National Health and Medical Research Council 1999, Oxford Centre for Evidence-based Medicine 2001). Esto enfatiza la preocupación general acerca de los sesgos

en los ENA y las dificultades para atribuir causalidad a los efectos observados. La fuerza de la evidencia proporcionada por una revisión sistemática de ENA es probable que dependa de vencer los retos establecidos en la Sección 13.7.1. La capacidad de vencer estos retos variará con el contexto de atención sanitaria y el resultado. En algunos contextos es probable que aparezcan pocos factores de confusión. Por ejemplo, es posible que se conozca muy poca información pronóstica cuando se vacuna a los niños, lo que limita los posibles factores de confusión (Jefferson 2005).

El hecho de que el debate concluya que existe la necesidad de ensayos aleatorizados o que la evidencia obtenida de ENA es adecuada para tomar decisiones informadas dependerá del coste de la incertidumbre que surge con el uso de diseños de estudios potencialmente sesgados, y del valor colectivo de los efectos observados. Este valor puede depender del contexto de atención sanitario más amplio. Puede que no sea posible incluir las evaluaciones del valor dentro de una revisión, lo que se puede hacer evidente sólo como parte del debate más amplio que sigue a la publicación.

Por ejemplo, ¿las pruebas provenientes de los ENA acerca de un efecto adverso grave poco frecuente son adecuadas para decidir que una intervención no se debe utilizar? La evidencia es incierta (debido a la falta de ensayos aleatorizados) pero el valor de conocer que existe la posibilidad de un daño potencialmente grave es considerable, y se puede valorar como suficiente para retirar la intervención. (Se debe señalar que la valoración acerca de retirar una intervención puede depender de si es posible obtener beneficios equivalentes de cualquier otra sin estos riesgos; de no ser así, la intervención aún se puede ofrecer pero con una revelación completa de sus daños potenciales). Cuando la evidencia de beneficio no se base en ensayos aleatorizados y por lo tanto sea equívoca, el valor adicional de una revisión sistemática de ENA sobre daños puede ser aún mayor.

Por otra parte, es posible que la evidencia de un beneficio pequeño de una intervención novedosa proporcionada por una revisión sistemática de ENA no sea suficiente para que los que toman las decisiones recomienden la implementación amplia, debido a la incertidumbre de la evidencia y los costes significativos que surgen de proporcionar la intervención. En estas circunstancias es probable que quienes toman las decisiones concluyan que se deberían realizar ensayos aleatorizados si es practicable y si es probable que la inversión en el ensayo se revierta en el futuro.

El uso del esquema GRADE para la evaluación de la calidad de un cuerpo de evidencia se recomienda en las tablas “Resumen de los hallazgos” de las revisiones Cochrane, y se resume en el Capítulo 12 (Sección 12.2). Hay cuatro niveles de calidad: “alta”, “moderada”, “baja” y “muy baja”. Un grupo de estudios que se puede categorizar de manera cruda como ensayos aleatorizados comienza con el nivel más alto, y pueden disminuir de nivel debido a las limitaciones del estudio (riesgo de sesgo), la falta de direccionalidad de las pruebas, la heterogeneidad, la imprecisión o el sesgo de publicación. Los grupos de estudios observacionales comienzan con un nivel “bajo” y pueden aumentar su grado de clasificación debido a una magnitud grande del efecto, la ausencia de preocupación acerca de los factores de confusión o a un gradiente dosis-respuesta. Los revisores deberán valorar si las pruebas de los ENA se deberían pasar a un nivel superior desde uno menor o posiblemente (p.ej. en el caso de los ensayos cuasialeatorizados) disminuir desde un nivel más alto.

13.7.3 Recomendaciones para los posibles revisores

La realización de una revisión sistemática de ENA es mucho más difícil que la realización de una revisión sistemática de ensayos aleatorizados. Es probable que sea necesario tomar decisiones complejas que requieren asesoría metodológica o epidemiológica experta en cada estadio de la revisión. Los posibles revisores deberían, por lo tanto, tratar de colaborar con epidemiólogos o metodólogos, independientemente de si una revisión tiene

como objetivo investigar daños o beneficios, resultados a corto o a largo plazo o eventos frecuentes o poco frecuentes.

Los profesionales de la atención sanitaria están deseosos de involucrarse en la realización de revisiones de ENA en áreas donde hay pocos o ningún ensayo aleatorizado porque desean mejorar la base de la evidencia en sus áreas de especialidad (la motivación para la mayoría de los revisores Cochrane). Los metodólogos están deseosos de más revisiones sistemáticas de ENA que proporcionen información en las variadas áreas de incertidumbre en la metodología señaladas en estos capítulos. Sin embargo, los profesionales de la atención sanitaria también deberían reconocer que (a) es probable que los recursos requeridos para hacer una revisión sistemática de ENA sean mucho mayores que para una revisión sistemática de ensayos aleatorizados y (b) es probable que las conclusiones sean mucho más débiles y que puedan hacer contribuciones relativamente pequeñas al tema. Por lo tanto, los revisores y editores de los GCR deberían decidir en un estadio temprano si la inversión de recursos puede justificarse por la prioridad de la pregunta de investigación.

La unión del equipo necesario de profesionales de la salud y metodólogos puede ser más fácil para revisiones sistemáticas de ENA que calculen los efectos de una intervención sobre resultados adversos poco frecuentes y a largo plazo, por ejemplo, cuando se consideran los efectos secundarios de los fármacos. Sin embargo, estas revisiones pueden requerir el aporte de revisores especialistas adicionales, por ejemplo, con experiencia farmacológica relevante. Existe una necesidad imperiosa en muchas afecciones de salud de complementar las revisiones sistemáticas tradicionales de ensayos aleatorizados de efectividad con revisiones sistemáticas de efectos adversos (no planificados). Es probable que estas revisiones sistemáticas necesiten incluir ENA.

13.8 Información del capítulo

Autores: Barnaby C Reeves, Jonathan J Deeks, Julian PT Higgins y George A Wells del Grupo Cochrane de Métodos de Estudios No Aleatorizados.

La versión en inglés de este capítulo se debería citar como: Reeves BC, Deeks JJ, Higgins JPT, Wells GA. Chapter 13: Including non-randomized studies. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: Se agradece a Ole Olsen, Peter Gøtzsche, Angela Harden, Mustafa Soomro, Guido Schwarzer y Bev Shea por colaborar con las primeras versiones de diferentes secciones. También se desea agradecer a Laurent Audigé, Duncan Saunders, Alex Sutton, Helen Thomas y Gro Jamtved por sus comentarios sobre las versiones previas.

Recuadro 13.8.a: El Grupo Cochrane de Métodos de Estudios No Aleatorizados

El Grupo de Métodos de Estudios No Aleatorizados (GMENA) de la Colaboración Cochrane asesora al Grupo Directivo para establecer políticas y formular recomendaciones acerca de la inclusión de estudios no aleatorizados (ENA) de la efectividad de intervenciones en la atención sanitaria en las revisiones Cochrane. La afiliación a este grupo está abierta a cualquiera que desee contribuir activamente al trabajo del grupo. El trabajo del grupo es principalmente metodológico, en lugar de estar centrado en determinadas intervenciones de la atención sanitaria.

Las actividades de los miembros del GMENA incluyen:

- Desarrollar guías para ayudar a decidir cuándo incluir datos no aleatorizados en las revisiones Cochrane.
- Realizar investigaciones metodológicas con el uso de estudios no aleatorizados, incluidos métodos de búsqueda, evaluación de la calidad, metanálisis, deficiencias y uso inadecuado.
- Realizar investigaciones empíricas para comparar el sesgo en las revisiones sistemáticas que utilicen estudios aleatorizados y no aleatorizados e identificar las condiciones bajo las cuales los estudios aleatorizados y no aleatorizados han dado lugar a conclusiones similares y las situaciones en las cuales las conclusiones han sido claramente contradictorias.
- Recopilar ejemplos de preguntas de la atención sanitaria que (a) se han estudiado mediante estudios no aleatorizados y ensayos aleatorizados, y (b) no se han estudiado adecuadamente (o que no se han estudiado durante un largo período) a través de ensayos aleatorizados.
- Proporcionar formación en el Coloquio Cochrane anual.

13.9 Referencias

Audige 2004
Audige L, Bhandari M, Griffin D, Middleton P, Reeves BC. Systematic reviews of nonrandomized clinical studies in the orthopaedic literature. *Clinical Orthopaedics and Related Research* 2004: 249-257.

Chan 2004
Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004; 291: 2457-2465.

Concato 1992
Concato J, Horwitz RI, Feinstein AR, Elmore JG, Schiff SF. Problems of comorbidity in mortality after prostatectomy. *JAMA* 1992; 267: 1077-1082.

Deeks 2003
Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG. Evaluating non-randomised intervention studies. *Health Technology Assessment* 2003; 7: 27.

Doll 1993
Doll R. Doing more good than harm: The evaluation of health care interventions: Summation of the conference. *Annals of the New York Academy of Sciences* 1993; 703: 310-313.

Downs 1998
Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health* 1998; 52: 377-384.

Eccles 1996
Eccles M, Clapp Z, Grimshaw J, Adams PC, Higgins B, Purves I, Russel I. North of England evidence based guidelines development project: methods of guideline development. *BMJ* 1996; 312: 760-762.

Fraser 2006
Fraser C, Murray A, Burr J. Identifying observational studies of surgical interventions in MEDLINE and EMBASE. *BMC Medical Research Methodology* 2006; 6: 41.

Furlan 2006
Furlan AD, Irvin E, Bombardier C. Limited search strategies were effective in finding relevant nonrandomized studies. *Journal of Clinical Epidemiology* 2006; 59: 1303-1311.

Glasziou 2007
Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007; 334: 349-351.

Golder 2006a
Golder S, Loke Y, McIntosh HM. Room for improvement? A survey of the methods used in systematic reviews of adverse effects. *BMC Medical Research Methodology* 2006; 6: 3.

Golder 2006b
Golder S, McIntosh HM, Duffy S, Glanville J, Centre for Reviews and Dissemination and UK Cochrane Centre Search Filters Design Group. Developing efficient search strategies to

identify reports of adverse effects in MEDLINE and EMBASE. *Health Information and Libraries Journal* 2006; 23: 3-12.

Golder 2006c
Golder S, McIntosh HM, Loke Y. Identifying systematic reviews of the adverse effects of health care interventions. *BMC Medical Research Methodology* 2006; 6: 22.

Greenland 2004
Greenland S. Interval estimation by simulation as an alternative to and extension of confidence intervals. *International Journal of Epidemiology* 2004; 33: 1389-1397.

Grobbee 1997
Grobbee DE, Hoes AW. Confounding and indication for treatment in evaluation of drug treatment for hypertension. *BMJ* 1997; 315: 1151-1154.

Henry 2001
Henry D, Moxey A, O'Connell D. Agreement between randomized and non-randomized studies: the effects of bias and confounding. *9th Cochrane Colloquium*, Lyon (France), 2001.

Higgins 2003
Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557-560.

Jefferson 2005
Jefferson T, Smith S, Demicheli V, Harnden A, Rivetti A, Di Pietrantonj C. Assessment of the efficacy and effectiveness of influenza vaccines in healthy children: systematic review. *The Lancet* 2005; 365: 773-780.

King 2005
King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, Sibbald B, Lai R. Impact of participant and physician intervention preferences on randomized trials: a systematic review. *JAMA* 2005; 293: 1089-1099.

Kwan 2004
Kwan J, Sandercock P. In-hospital care pathways for stroke. *Cochrane Database of Systematic Reviews* 2004, Issue 2. Art No: CD002924.

MacLehose 2000
MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AM. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technology Assessment* 2000; 4: 1-154.

Moher 2001
Moher D, Schulz KF, Altman DG. The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *The Lancet* 2001; 357: 1191-1194. (Available from www.consort-statement.org).

National Health and Medical Research Council 1999
National Health and Medical Research Council. *A guide to the development, implementation and evaluation of clinical practice guidelines [Endorsed 16 November 1998]*. Canberra (Australia): Commonwealth of Australia, 1999.

Oxford Centre for Evidence-based Medicine 2001
Oxford Centre for Evidence-based Medicine. Levels of Evidence [May 2001]. Available from: http://www.cebm.net/index.aspx?o=1047 (accessed 1 January 2008).

Peto 1995
Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. *Journal of Clinical Epidemiology* 1995; 48: 23-40.

Petticrew 2001
Petticrew M. Systematic reviews from astronomy to zoology: myths and misconceptions. *BMJ* 2001; 322: 98-101.

Pocock 2004
Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, Kalish LA, Kasten LE, McCormack VA. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ* 2004; 329: 883.

Reeves 2004
Reeves BC, Gaus W. Guidelines for reporting non-randomised studies. *Forschende Komplementärmedizin und klassische Naturheilkunde* 2004; 11 Suppl 1: 46-52.

Reeves 2006
Reeves BC. Parachute approach to evidence based medicine: as obvious as ABC. *BMJ* 2006; 333: 807-808.

Reeves 2007
Reeves BC, Langham J, Lindsay KW, Molyneux AJ, Browne JP, Copley L, Shaw D, Gholkar A, Kirkpatrick PJ. Findings of the International Subarachnoid Aneurysm Trial and the National Study of Subarachnoid Haemorrhage in context. *British Journal of Neurosurgery* 2007; 21: 318-23.

Rothman 1986
Rothman KJ. *Modern Epidemiology*. Boston (MA): Little, Brown & Company, 1986.
Shadish 2002
Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston (MA): Houghton Mifflin, 2002.

Siegfried 2003
Siegfried N, Muller M, Volmink J, Deeks J, Egger M, Low N, Weiss H, Walker S, Williamson P. Male circumcision for prevention of heterosexual acquisition of HIV in men. *Cochrane Database of Systematic Reviews* 2003, Issue 3. Art No: CD003362.

Taggart 2001
Taggart DP, D'Amico R, Altman DG. Effect of arterial revascularisation on survival: a systematic review of studies comparing bilateral and single internal mammary arteries. *The Lancet* 2001; 358: 870-875.

Vandenbroucke 2007
Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Medicine* 2007; 4: e297.

von Elm 2007
von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *PLoS Medicine* 2007; 4: e296.

Wells 2008
Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses.

Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm (accessed 1 January 2008).

Wieland 2005
Wieland S, Dickersin K. Selective exposure reporting and Medline indexing limited the search sensitivity for observational studies of the adverse effects of oral contraceptives. *Journal of Clinical Epidemiology* 2005; 58: 560-567.

ÍNDICE

CAPÍTULO 14: EFECTOS ADVERSOS

PUNTOS CLAVE

14.1 INTRODUCCIÓN

14.1.1 La necesidad de considerar los efectos adversos
14.1.2 Conceptos y terminología
14.1.3 Cuándo es más importante considerar los efectos adversos
Tabla 14.1.a: Contextos y ejemplos que justifican el examen detallado de los efectos adversos

14.2 ALCANCE DE UNA REVISIÓN QUE ABORDA LOS EFECTOS ADVERSOS

14.2.1 Métodos idénticos para los efectos beneficiosos y adversos
14.2.2 Métodos diferentes para los efectos beneficiosos y adversos
14.2.3 Revisión separada para los efectos adversos

14.3 ELECCIÓN DE LOS EFECTOS ADVERSOS A INCLUIR

14.3.1 Enfoque estrecho frente a amplio
14.3.2 Retiro o abandono como una medida de resultado para los efectos adversos

14.4 TIPOS DE ESTUDIOS

14.5 MÉTODOS DE BÚSQUEDA PARA LOS EFECTOS ADVERSOS

14.5.1 Fuentes de información sobre los efectos adversos de los fármacos
14.5.2 Estrategia de búsqueda para los efectos adversos

Introducción

14.5.2.1 Búsqueda en las bases de datos electrónicas para los efectos adversos mediante los términos indexados
14.5.2.2 Búsqueda en bases de datos electrónicas para efectos adversos mediante términos de texto libre

14.6 EVALUACIÓN DEL RIESGO DE SESGO PARA LOS EFECTOS ADVERSOS

14.6.1 Ensayos clínicos
14.6.2 Estudios de casos y controles y estudios de cohortes
14.6.3. Informe de casos

14.7 INFORMACIÓN DEL CAPÍTULO

Recuadro 14.7.a: El Grupo Cochrane de Métodos de Efectos Adversos

14.8 REFERENCIAS

CAPÍTULO 14: EFECTOS ADVERSOS

Autores: Yoon K Loke, Deirdre Price y Andrew Herxheimer del Grupo Cochrane de Métodos de Efectos Adversos.

Puntos clave

- Para lograr una perspectiva equilibrada, todas las revisiones deberían tratar de considerar los aspectos adversos de las intervenciones.
- Un análisis detallado de los efectos adversos es particularmente relevante cuando la evidencia sobre la posibilidad de daño tiene una influencia importante sobre el tratamiento o la política de decisiones.
- Las intervenciones pueden tener muchos efectos adversos diferentes, y las revisiones necesitan centrarse detalladamente en algunos importantes, junto con un resumen amplio y más general de otros potenciales efectos adversos.
- Como con frecuencia los datos de los efectos adversos se tratan de forma menos rigurosa que los resultados principales de un estudio, la intensidad de la monitorización de los efectos adversos y la claridad de su descripción se deberían revisar cuidadosamente.
- Los datos sobre los efectos adversos con frecuencia son escasos, pero la falta de información no significa que la intervención sea segura.

14.1 Introducción

14.1.1 La necesidad de considerar los efectos adversos

Cada intervención de atención sanitaria implica el riesgo, grande o pequeño, de efectos adversos o perjudiciales. Una revisión Cochrane que considere solamente los resultados favorables de las intervenciones que examina, sin evaluar también los efectos adversos, presentará un desequilibrio y hará que la intervención analizada sea más favorable de lo que debería ser. Esta fuente de sesgo, al igual que otras, se debería minimizar. En todas las revisiones los revisores deberían tratar de incluir alguna consideración sobre los aspectos adversos de las intervenciones.

Este capítulo aborda aspectos especiales relacionados con los efectos adversos en las revisiones Cochrane, con énfasis en las revisiones en las que los efectos adversos podrían analizarse mediante métodos que difieran de los de otros resultados. Aunque en principio los efectos adversos son más fiables cuando se evalúan a través de ensayos aleatorizados, en la práctica muchos de los eventos adversos son muy poco frecuentes o suceden a muy largo plazo para ser observados en ensayos aleatorizados, o puede ser que no se conocieran cuando los ensayos se planificaron. Una revisión Cochrane puede utilizar una de varias estrategias para analizar los efectos adversos, las cuales difieren en el grado en el cual los mismos métodos se utilizan para evaluar los efectos intencionados (beneficiosos) o no intencionados (beneficiosos o adversos). El presente capítulo se enfoca en los afectos adversos que habitualmente se consideran no intencionados (Miettinen 1983). Las diferentes estrategias para una revisión se comentan en la Sección 14.2.

14.1.2 Conceptos y terminología

En la atención sanitaria se utilizan muchos términos para describir los daños asociados con las intervenciones. Este hecho puede confundir a los revisores, particularmente debido a que con frecuencia los artículos publicados utilizan términos imprecisos e intercambiables. Algunos de los términos frecuentes relacionados incluyen “evento adverso” (un resultado desfavorable que ocurre durante o después del uso de un fármaco u otra intervención pero que no es causado necesariamente por éste), “efecto adverso” (un evento adverso para el cual la relación causal entre la intervención y el evento es al menos una posibilidad razonable), “reacción adversa al fármaco” (un afecto adverso específico de un fármaco), “efecto secundario” (cualquier efecto no intencionado, adverso o beneficioso, de un fármaco que ocurra a dosis utilizadas normalmente para el tratamiento) y “complicaciones” (eventos o efectos adversos posteriores a intervenciones quirúrgicas u otras intervenciones invasivas).

14.1.3 Cuándo es más importante considerar los efectos adversos

Los recursos dedicados a incluir los resultados adversos en las revisiones se deberían considerar en relación a la importancia de la revisión en sí misma. Si una intervención claramente no funciona o tiene un beneficio potencial pequeño y no se utiliza ampliamente, es posible que no merezca la pena dedicar recursos a una evaluación detallada de los efectos adversos. Por otro lado, un análisis detallado de los efectos adversos se podría justificar si la información sobre los daños potenciales resulta esencial en la guía de las decisiones de los médicos, los consumidores y los elaboradores de políticas.

La Tabla 14.1.a aporta ejemplos de las situaciones en las que el análisis de los efectos adversos tiene una función importante en las decisiones de tratamiento.

Tabla 14.1.a: Contextos y ejemplos que justifican el examen detallado de los efectos adversos

Cuando el margen entre beneficios y efectos adversos es estrecho	
El tratamiento es de un beneficio modesto o incierto, con una posibilidad importante de efectos adversos	▪ Aspirina para la prevención de los eventos cardiovasculares en un paciente saludable; aumento de la hemorragia. ▪ Antibióticos para la otitis media aguda en niños; riesgo de erupción cutánea y diarrea. ▪ Cardioversión urgente con corriente directa en pacientes con una fibrilación atrial nueva que están en estado cardiovascular estable; riesgo de accidente cerebrovascular debido a la cardioversión.
El tratamiento es potencialmente muy beneficioso, pero existen dudas importantes sobre su seguridad.	▪ Aspirina para pacientes con un accidente cerebrovascular, pero que tienen antecedentes de hemorragia gastrointestinal. ▪ Endoarteriectomía carotídea en pacientes de edad avanzada con cardiopatía isquémica que presentan un accidente cerebrovascular.
El tratamiento es potencialmente beneficioso a largo plazo o para la comunidad, pero no beneficia directamente de forma inmediata al individuo.	▪ Mejoría en la administración de una vacuna para promover la inmunidad de grupo, mientras se intenta mitigar los miedos sobre los rápidos efectos adversos neurológicos graves.
Cuando varios tratamiento eficaces difieren en sus perfiles de seguridad	
Los tratamientos tienen una eficacia equivalente pero tienen perfiles de seguridad diferentes.	▪ Fármacos antiepilépticos para mujeres en edad reproductiva con epilepsia. ▪ Un dispositivo nuevo para inyectar la insulina se considera que provoca menos dolor que el dispositivo existente.
El equilibrio entre los beneficios y los efectos adversos difiere significativamente, p.ej. la intervención más eficaz puede tener efectos adversos graves, mientras que la intervención menos efectiva es potencialmente segura.	▪ Fármaco que modifica la enfermedad en la artritis reumatoide erosiva, p.ej. uso de hidroxicloroquina (relativamente segura) o metrotexato (potencialmente más efectivo, pero menos seguro). ▪ Poliquimioterapia versus quimioterapia con un agente único secuencial para el cáncer de mama metastático.
Cuando los efectos adversos impiden que un paciente continúe con un tratamiento eficaz	
El tratamiento tiene un beneficio considerable pero los efectos adversos impiden el cumplimiento del paciente y se necesita evidencia para guiar el tratamiento adicional.	▪ Una intervención efectiva tiene efectos adversos bien reconocidos que pueden provocar dificultades para que el paciente continúe el tratamiento. Se necesita evidencia sobre si reducir la intensidad de la intervención (p.ej. disminución de la dosis o la duración) ayudaría a evitar los efectos adversos o si existe una estrategia de tratamiento que pueda prevenir los efectos adversos (p.ej. inhibidor de la bomba de protones para las úlceras pépticas provocadas por la aspirina).

14.2 Alcance de una revisión que aborda los efectos adversos

14.2.1 Métodos idénticos para los efectos beneficiosos y adversos

En esta sección y en las Secciones 14.2.2 y 14.2.3, se describen las estrategias amplias que se pueden utilizar en una revisión Cochrane para abordar los efectos adversos. La primera estrategia consiste en evaluar los efectos intencionados (beneficiosos) y no intencionados (adversos) juntos mediante la misma metodología y aplicar criterios de elegibilidad comunes (con respecto al tipo de estudios, tipo de participantes y tipo de intervenciones).

Este enfoque implica que se utilice una estrategia de búsqueda única. Un aspecto crítico es qué harán los revisores con los tres conjuntos de datos que podrían surgir:
(a) Estudios que describan los efectos beneficiosos y los efectos adversos de interés.
(b) Estudios que describan los efectos beneficiosos pero no los efectos adversos.
(c) Estudios que describan los efectos adversos, pero no los resultados beneficiosos de interés.

Los estudios del tipo (a) tienen la importante ventaja de que los beneficios y los efectos adversos se pueden comparar de forma directa, ya que los datos provienen de la misma población y ámbito. Además la evidencia de los beneficios y efectos adversos provienen de estudios de diseño y calidad similares. Sin embargo, los datos sobre los efectos adversos pueden ser muy limitados y en particular pueden estar restringidos a daños a corto plazo debido a la duración relativamente corta de los estudios incluidos.

La evaluación de los beneficios y los efectos adversos mediante alguna combinación de los tres tipos de estudios (en lugar de los [a] solos) aumentaría la cantidad de información disponible. Por ejemplo, los conjuntos de datos (a) y (b) se podrían utilizar para evaluar los efectos beneficiosos mientras que los de (a) y (c) se podrían utilizar para evaluar los efectos adversos. Sin embargo, como los estudios que abordan los efectos adversos difieren de los que abordan los efectos beneficiosos, los revisores deberían tener en cuenta que es difícil comparar directamente los beneficios y los efectos adversos.

14.2.2 Métodos diferentes para los efectos beneficiosos y adversos

La segunda estrategia consiste en utilizar criterios de elegibilidad diferentes para seleccionar los estudios que abordan efectos no intencionados (adversos) comparados con los estudios que abordan los efectos intencionados (beneficiosos).

Para evaluar resultados diferentes pueden ser necesarios diferentes tipos de estudios (Glasziou 2004). El uso de criterios de elegibilidad diferentes aborda específicamente el problema de que la mayoría de los estudios experimentales (como los ensayos aleatorizados) no son suficientes para evaluar los efectos adversos poco frecuentes, a largo plazo y no reconocidos previamente (ver Sección 14.4). Este enfoque permite una evaluación más rigurosa de los efectos adversos, pero consume más tiempo y recursos, lo que significa que con frecuencia los beneficios y los efectos adversos no se pueden comparar directamente. Aunque los ensayos aleatorizados tienen la ventaja de que la asignación de las intervenciones se realiza mediante un proceso de asignación al azar, los estudios no aleatorizados aplican mecanismos diferentes para asignar las intervenciones, y este aspecto se debería examinar durante la revisión.

14.2.3 Revisión separada para los efectos adversos

La tercera estrategia consiste en realizar una revisión separada de los efectos adversos solos. Esto podría ser apropiado para una intervención que se aplique en una variedad de enfermedades o afecciones, aun cuando su perfil de efectos adversos se podría esperar que fuera similar en diferentes poblaciones y ámbitos. Por ejemplo, la aspirina se utiliza en una amplia variedad de pacientes, como los que presentan un accidente cerebrovascular o enfermedad vascular periférica, y también en aquellos con arteriopatía coronaria. Típicamente los efectos principales de la aspirina sobre los resultados relevantes para estas afecciones

diferentes podrían ser abordados por separado en las revisiones Cochrane, pero los efectos adversos (como la hemorragia cerebral o intestinal) son suficientemente similares dentro de los grupos de enfermedades diferentes como para poder abordarlos juntos en una revisión independiente. De hecho, a menos de que existan ensayos en poblaciones combinadas, este aspecto sería difícil de abordar de otra manera.

De forma parecida, pueden existir datos limitados sobre los efectos adversos en una subpoblación, como los niños. Puede ser útil analizar todos los datos disponibles para esta subpoblación (por ej. efectos adversos de los inhibidores selectivos de la recaptación de serotonina en niños), incluso si los ensayos se dirigieron a diferentes enfermedades.

Los autores de las revisiones de efectos adversos solos deberían tener como objetivo proporcionar una adecuada referencia cruzada (preferiblemente mediante enlaces electrónicos) a revisiones relacionadas que traten los efectos intencionados de la intervención. Si al actualizar una revisión de eficacia se identifican nuevos aspectos de seguridad, la revisión de efectos adversos se debería actualizar tan pronto como sea posible.

14.3 Elección de los efectos adversos a incluir

14.3.1 Enfoque estrecho frente a amplio

La selección de los resultados adversos a incluir en una revisión puede ser difícil. Es posible que los efectos adversos específicos asociados con una intervención se conozcan antes de la revisión, mientras que otros no. Podría resultar incierto conocer de antemano los efectos que podrían ser más relevantes para la revisión. Las siguientes estrategias generales se pueden utilizar según la pregunta del estudio y el contexto terapéutico o preventivo.

Enfoque estrecho

Un análisis detallado de uno o dos efectos adversos conocidos o algunos de los efectos adversos más graves que son de interés especial para los pacientes y los profesionales de la salud.
Ventajas: Enfoque más fácil, especialmente con respecto a la obtención de los datos. Puede centrarse en los efectos adversos importantes y alcanzar una conclusión significativa sobre aspectos que tienen una repercusión importante sobre la decisión de tratamiento (McIntosh 2004).
Desventajas: El enfoque puede ser demasiado estrecho. El método sólo es realmente adecuado para los eventos adversos que se conocen previamente.

Enfoque amplio

Para detectar una variedad de efectos adversos, conocidos o no conocidos previamente.
Ventajas: Amplia cobertura, y puede evaluar efectos adversos nuevos que es posible que no se conozcan previamente.
Desventajas: Posiblemente un gran volumen de trabajo con determinadas dificultades en la recogida de datos. Algunos investigadores encuentran que las evaluaciones amplias no específicas necesitan muchos recursos y muestran muy poca información con respecto al esfuerzo invertido (McIntosh 2004). Estos investigadores también señalan que los efectos adversos no reconocidos previamente se pueden detectar mejor mediante una monitorización primaria, en lugar de en una revisión sistemática.
Para abordar los efectos adversos de una forma más organizada, los revisores pueden seleccionar un enfoque estrecho en algunas de las áreas siguientes:

- Los cinco a diez efectos adversos más frecuentes.
- Todos los efectos adversos que el paciente o el médico consideren que son graves.

- Por categoría, por ejemplo:
  - Diagnosticado por resultados de laboratorio (p.ej. hipopotasemia); o
  - Síntomas notificados por el paciente (p.ej. dolor).

14.3.2 Retiro o abandono como una medida de resultado para los efectos adversos

Con frecuencia el retiro o abandono se utiliza como una medida de resultado en los publicaciones de los ensayos. Los revisores no deberían interpretar tales datos como marcadores sustitutos para la seguridad y la tolerancia debido a la posibilidad de sesgo:

- La atribución de el/los motivo/s para la interrupción es compleja y puede ser debida a efectos secundarios leves pero irritantes, toxicidad, falta de eficacia, motivos no médicos o una combinación de causas (Ioannidis 2004).
- Las presiones sobre los pacientes e investigadores según las condiciones del ensayo para mantener bajo el número de retiros y abandonos puede provocar tasas que no reflejen la ocurrencia de eventos adversos en la población de estudio.
- Con frecuencia la falta de cegamiento de la asignación a la intervención precede la decisión de retirarse. Lo anterior puede provocar una sobrestimación del efecto de la intervención sobre el retiro de los pacientes. Por ejemplo, es menos probable que los síntomas de los pacientes de la rama placebo provoquen la interrupción. Por el contrario, los pacientes del grupo de intervención activa que se quejan de síntomas que indican efectos adversos puede ser que se retiren más fácilmente.

14.4 Tipos de estudios

La mayoría de las revisiones Cochrane se centran en los ensayos aleatorizados, los cuales proporcionan estimaciones más fiables del efecto. Sin embargo, en los ensayos clínicos es poco probable que se observen eventos adversos poco frecuentes o efectos adversos a largo plazo, y una investigación rigurosa puede requerir la inclusión de estudios de cohortes, estudios de casos y controles e incluso informes de casos o series de casos. En particular, es probable que se elijan específicamente las estrategias esbozadas en las Secciones 14.2.2 y 14.2.3 debido a que para abordar los efectos adversos se incluyen diferentes diseños de estudios. Para una discusión más detallada de los aspectos en la inclusión de estudios no aleatorizados (incluidos los estudios de casos y controles y los estudios de cohortes) en una revisión Cochrane, ver Capítulo 13 (Sección 13.2). Algunos aspectos a considerar en la inclusión de los informes de casos aparecen en la Sección 14.6.3.

14.5 Métodos de búsqueda para los efectos adversos

14.5.1 Fuentes de información sobre los efectos adversos de los fármacos

Además de las fuentes de pruebas habituales, descritas en el Capítulo 6, los revisores que planifican una búsqueda exhaustiva de los efectos adversos de un fármaco deberían considerar verificar las siguientes fuentes:

- Libros de referencia estándar sobre efectos adversos como Meyler's Side Effects of Drugs, los Side Effects of Drugs Annuals (SEDA), Martindale: The Complete Drug Reference, Davies Textbook of Adverse Drug Reactions y los artículos que ellos resumen.
- Las autoridades reguladoras pueden emitir alertas de seguridad para varios productos comerciales según la información que el fabricante les envía (la cual no se ha publicado ni está disponible en lugar alguno). Ejemplos de boletines de seguridad se pueden encontrar:
  - en el RU: *Current Problems in Pharmacovigilance* (www.mhra.gov.uk);
  - en Australia: el *Australian Adverse Drug Reactions Bulletin* (www.tga.gov.au/adr/aadrb.htm);

en los *European Public Assessment Reports from the European Medicines Evaluation Agency* (www.emea.eu);
en los EE.UU.: *Food and Drug Administration FDA Medwatch* (www.fda.gov/medwatch).

- Las bases de datos especializadas en información de fármacos como las bases de datos de texto completo (p.ej *Pharmanewsfeed* and *Iowa Drug Information Service* [IDIS]), las bases de datos bibliográficas (p.ej. *Derwent Drug File*, *TOXLINE*, *Pharmline*) y las bases de datos de resumen de referencias (p.ej. *Drugdex*, *XPhram*). Sin embargo, los revisores deberán considerar los costes de suscripción a estas bases de datos especializadas, particularmente cuando su utilidad o rendimiento adicional aún tienen que ser evaluados formalmente en el ámbito de las revisiones sistemáticas.

Los revisores también pueden recurrir (habitualmente mediante el pago de una cuota) al *WHO Uppsala Monitoring Centre* (UMC; www.who-umc.org) para búsquedas especiales en sus bases de datos de información espontánea (*Vigibase*); lo que se realizó, por ejemplo, para una revisión Cochrane sobre melatonina (Herxheimer 2002). Sin embargo se encontró que el rango de orden de la mayoría de los efectos adversos más frecuentes descritos para un determinado fármaco en la base de datos UMC difería de los datos obtenidos de un metanálisis de ensayos aleatorizados doble ciego (Loke 2004): los datos de la UMC sobre amiodarona mostraron que fueron más frecuentes los problemas del tiroides y las reacciones cutáneas estuvieron en segundo lugar, mientras que el metanálisis mostró que los problemas cardíacos fueron más frecuentes, seguidos por los trastornos del tiroides.

Los datos de vigilancia primaria (en forma de casos clínicos espontáneos) también están disponibles libremente mediante los sitios web de las autoridades reguladoras en Canadá, EE.UU., y los Países Bajos. Sin embargo, el formato de la información varía considerablemente, y la interpretación y análisis de estas bases de datos requiere habilidades especiales (ver también Sección 14.6.3).

14.5.2 Estrategia de búsqueda para los efectos adversos

Introducción

Aún es necesario establecer la estrategia de búsqueda óptima para identificar específicamente los informes de efectos adversos (Golder 2006). Se pueden utilizar dos enfoques principales: la búsqueda mediante términos indexados y la búsqueda de texto libre. Ambos tienen limitaciones y es aconsejable combinarlos para maximizar la sensibilidad (la probabilidad de no perder estudios que pudieran ser relevantes). Es probable que el desarrollo de la estrategia de búsqueda requiera algunas iteraciones. Por ejemplo, puede que sea necesario repetir la búsqueda electrónica para incorporar términos indexados adicionales, subtítulos y términos de texto libre obtenidos de los términos utilizados para indicar y describir los estudios identificados inicialmente como relevantes. Para decidir qué combinación de términos utilizar, los revisores deberán equilibrar la exhaustividad (sensibilidad) con la precisión. A continuación se tratan algunas consideraciones en el uso de los términos indizados y los términos de texto libre.

14.5.2.1 Búsqueda en las bases de datos electrónicas para los efectos adversos mediante los términos indexados

Los términos indexados (también llamados vocabulario controlado o términos de tesauro) como los Medical Subject Headings (MeSH) en MEDLINE y EMTREE en EMBASE se asignan a registros en las bases de datos electrónicas para describir los estudios. MEDLINE y EMBASE emplean algunos términos indexados útiles para los efectos adversos: incluyen *DRUG TOXICITY/* y *ADVERSE DRUG REACTION SYSTEMS* en MEDLINE y *DRUG TOXICITY/* y *ADVERSE DRUG REACTION/* en EMBASE. Sin embargo, la forma más útil para buscar los efectos adversos es mediante los subtítulos (Golder 2006). Los subtítulos se pueden adjuntar a los términos indexados para describir aspectos específicos, por ejemplo, “efectos secundarios” de los fármacos o “complicaciones” de la cirugía, o se pueden utilizar en las búsquedas para adjuntarlos a cualquier término indexado (subtítulos flotantes). Los subtítulos utilizados para

denotar los datos sobre los efectos adversos difieren en las bases de datos importantes MEDLINE y EMBASE, por ejemplo:

Aspirin/adverse effects (MEDLINE)
Acetylsalicylic-acid/adverse-drug-reaction (EMBASE)

En el ejemplo anterior *Aspirin* es el término MeSH y *adverse effects* es el subtítulo; *Acetylsalicylic-acid* es el término EMTREE y *adverse-drug-reaction* es el subtítulo.

Dentro de una base de datos los estudios pueden estar (i) indexados bajo el nombre de la intervención junto con un subtítulo para denotar los efectos adversos ocurridos, por ejemplo, *Aspirin/adverse effects* o *Mastectomy/complications*; o (ii) el evento adverso en sí mismo puede ser indexado, junto con la naturaleza de la intervención, por ejemplo, *Gastrointestinal Hemorrhage/ and Aspirin/* , o *Lymphedema/ and Surgery/*; u (iii) ocasionalmente, un artículo puede estar indexado sólo bajo el evento adverso, por ejemplo, *Hemorrhage/chemically-induced*.

Por lo tanto, no se puede confiar en un solo término de búsqueda índice o subtítulo para identificar todos los datos sobre los efectos adversos, pero una combinación de términos indexados y subtitulados es útil para detectar informes de efectos adversos principales que es probable que los indexadores consideren significativos (Derry 2001).

Los subtítulos que se pueden utilizar con la intervención o con todas las intervenciones (flotantes) y que pueden ser útiles en MEDLINE son:

/adverse effects (nótese que si se utiliza este subtítulo incluirá los subtítulos */poisoning and /toxicity*)
/poisoning
/toxicity
/contraindications

Los subtítulos que se pueden utilizar con el resultado adverso o con todos los resultados (flotantes) y que pueden ser útiles en MEDLINE son:

/chemically induced
/complications

Los subtítulos que se pueden utilizar con la intervención o con todas las intervenciones (flotantes) y que pueden ser útiles en EMBASE son:

/adverse drug reaction
/drug toxicity

Los subtítulos que se pueden utilizar con el resultado adverso o con todos los resultados (flotantes) y que pueden probar su utilidad en EMBASE son:

/complication
/side effect

14.5.2.2 Búsqueda en bases de datos electrónicas para efectos adversos mediante términos de texto libre

Los autores utilizan los términos de texto libre (también llamados palabras de texto) en el título y el resumen de sus estudios cuando se publican como artículos de revistas; por lo que es posible buscar estos términos en el título y el resumen de los registros electrónicos en las bases de datos. Hay dos problemas importantes que limitan la utilidad de la búsqueda de texto libre:

1. El amplio rango de términos que utilizan los autores para describir los efectos adversos en sentido general (toxicidad, efecto secundario, efectos adversos) y de forma más específica (por ejemplo, letargia, cansancio, malestar pueden utilizarse como sinónimos).

2. La búsqueda de texto libre no detecta efectos adversos no mencionados en el título o el resumen del estudio y, por lo tanto, no se incluyen en el registro electrónico (incluso aunque el informe completo los describa) (Derry 2001).

Una búsqueda altamente sensible con texto libre debería incorporar la posible amplia variedad de términos sinónimos, a la vez que tendría en cuenta diferentes convenciones en el deletreo y las variaciones en el final de los términos a incluir, por ejemplo términos en singular y en plural. Luego se debe combinar con términos de texto libre que incluyan la intervención de interés, por ejemplo:
(aspirin or acetylsalicylic acid) and (adverse or side or hemorrhage or haemorrhage or bleed or bleeding or blood loss).

14.6 Evaluación del riesgo de sesgo para los efectos adversos

14.6.1 Ensayos clínicos

Aunque el consejo general es evaluar el riesgo de sesgo en los ensayos clínicos como se describe en el Capítulo 8, los revisores también deberían considerar otros factores específicos que pueden tener una influencia mayor sobre los datos de efectos adversos. Las áreas de mayor preocupación incluyen los métodos para monitorizar y detectar los efectos adversos, los conflictos de interés (Jüni 2004), el informe selectivo de resultado (Chan 2004) y el cegamiento (Schulz 2002).

La medida de resultado primaria de una intervención puede haberse estudiado en un ensayo aleatorizado con un ocultamiento y enmascaramiento adecuados y controlado con placebo. Por el contrario, es posible que los datos de los efectos adversos se obtengan de forma retrospectiva, por ejemplo, a través de un cuestionario al final del estudio enviado solamente a quienes se conoce que recibieron la intervención activa. Aunque se puede asignar un riesgo de sesgo bajo a los resultados primarios, es posible que la manera en la cual los efectos perjudiciales de las intervenciones se monitorizan no permita una clasificación similar. La herramienta recomendada para el riesgo de sesgo, implementada en RevMan, permite evaluaciones diferentes del cegamiento y de los datos de resultado incompletos para cada resultado, o para una clase de resultados como los definió el revisor.

Se sabe que los métodos utilizados para monitorizar o detectar los efectos adversos tienen una influencia importante en las frecuencias de los efectos adversos: los estudios en los cuales los efectos adversos se investigan de forma cuidadosa notificarán una mayor frecuencia de efectos adversos que los estudios en los cuales se investigan con menos cuidado. Por ejemplo, en un grupo de pacientes hipertensos, la monitorización pasiva basada en los informes espontáneos produjo tasas del 16%, mientras que la monitorización activa mediante cuestionarios específicos encontró una tasa del 62% (Olsen 1999). Como diferentes métodos para monitorizar los efectos adversos pueden producir diferentes resultados, puede ser difícil comparar los estudios y sería inútil realizar un metanálisis formal (Edwards 1999). También se deberían señalar la duración y la frecuencia de la monitorización.

Es posible que los estudios con un seguimiento limitado o con una monitorización poco frecuente no sean fiables para detectar los efectos adversos; la falta de información no se debería interpretar como indicadora de que la intervención es segura. Por el contrario, los estudios con un seguimiento riguroso y una vigilancia activa de efectos adversos predefinidos pueden ser capaces de generar pruebas de que la intervención tiene realmente pocos efectos adversos.

Finalmente, es probable que el tiempo que lleva implementada una intervención y la evolución de su uso se relacionen con el tipo de eventos adversos detectados y su número. Esto es

evidente para los efectos a largo plazo como la carcinogenia, pero también debido a que algunas intervenciones, como las quirúrgicas, cambian de manera más o menos sutil con el tiempo.

Ejemplos de preguntas potencialmente útiles a considerar cuando se evalúa la calidad de la evidencia sobre los efectos adversos son:

En la realización:
- ¿Se proporcionan las definiciones de los efectos adversos descritos?
- ¿Se describieron los métodos utilizados para la monitorización de los efectos adversos? Uso de monitorización prospectiva o habitual, informe espontáneo, lista de verificación del paciente, cuestionario o diario, encuesta sistemática de los pacientes.

En el informe:
- ¿Se excluyeron algunos pacientes del análisis de los efectos adversos?
- ¿El informe proporciona datos numéricos por grupo de intervención?
- ¿Qué categorías de efectos adversos describieron los investigadores?

14.6.2 Estudios de casos y controles y estudios de cohortes

Aunque el estudio de los efectos beneficiosos casi siempre necesita de ensayos aleatorizados, a menudo los efectos adversos del tratamiento se pueden investigar de manera efectiva en estudios no aleatorizados (Miettinen 1983). Vandenbroucke ha señalado que los estudios observacionales de efectos adversos de investigaciones médicas ofrecen algunas de las mejores oportunidades para los estudios observacionales no sesgados (Vandenbroucke 2004). Esta idea fue empíricamente verificada en una comparación de estudios aleatorizados y observacionales de efectos adversos, la cual encontró que las estimaciones de riesgo de los estudios observacionales fueron menores (Papanikolau 2006). En algunas circunstancias en las que los estudios observacionales mostraron riesgos marcadamente mayores se reflejó mejor la atención real de los pacientes (Vandenbroucke 2006). Como cualquier otro estudio, los estudios de casos y controles y los estudios de cohortes son potencialmente susceptibles al sesgo, por lo que se debería discutir críticamente cualquier limitación de los datos. Ver Capítulo 13 (Sección 13.5) para una discusión adicional de la evaluación del riesgo de sesgo en estos estudios. Jick ha descrito la taxonomía del tipo de estudio que es más probable que detecte un efecto adverso, así como el tipo de estudio necesario para la verificación (Jick 1977).

14.6.3 Informe de casos

Los informes de casos de eventos adversos se encuentran ampliamente en la bibliografía publicada, y también son recopilados por agencias reguladoras. Existen problemas metodológicos específicos con la evaluación de tales informes de casos. Los revisores que estén potencialmente interesados en tales datos necesitan considerar los siguientes aspectos.

¿Los informes tienen un valor predictivo adecuado?

Los informes anecdóticos pueden convertirse en falsas alarmas para investigaciones posteriores, en lugar de indicadores verdaderos del vínculo entre la intervención y el efecto adverso. Aunque un estudio ha señalado que tres cuartas partes de una colección de informes de casos anecdóticos desde 1963 fueron correctos (Venning 1982), una encuesta sistemática más reciente de 63 supuestas reacciones adversas encontró que la mayoría (52 de 63; 82,5%) aún no se habían evaluado en más detalle (Loke 2006). Los datos de estudios controlados que apoyan el vínculo postulado entre el fármaco y el evento adverso sólo estuvieron disponibles en tres casos, mientras que en dos casos los estudios controlados no pudieron confirmarlo. Sin embargo, las hojas de información del producto o las monografías del fármaco se pueden haber

modificado para incluir las listas de estos eventos adversos. Por lo tanto, no es fácil decir si un informe de caso es una alerta real o una falsa alarma. Los informes de casos todavía son la piedra angular de los efectos adversos nuevos (Stricker 2004). En el pasado y en la actualidad la retirada de los fármacos del mercado se basa fundamentalmente en los informes de casos y las series de casos (Venning 1983, Arnaiz 2001). La retirada de un fármaco del mercado debido a un efecto dramático no requiere grupos control formales (Glasziou 2007).

Determinación de la causalidad

Habitualmente existe incertidumbre en si la intervención provocó el evento adverso (particularmente en pacientes que reciben una amplia variedad de tratamientos). Los revisores deberían decidir si es probable que la intervención tenga un papel causal, o si la ocurrencia del evento adverso durante el período de intervención fue solamente una coincidencia. Sin embargo, es posible que dos revisores independientes no hagan la misma valoración del mismo informe de casos. Varios estudios han evaluado las respuestas de los revisores a los que se les solicitó que evaluaran informes de eventos adversos. En un estudio se obtuvo un acuerdo completo en sólo el 35% de las veces entre dos observadores que utilizaron los criterios de causalidad en un algoritmo para evaluar supuestas reacciones adversas (Lanctot 1995). En otro estudio tres farmacólogos clínicos que evaluaron 500 informes de supuestas reacciones no pudieron coincidir en el fármaco culpable en el 36% de los casos (Koch-Weser 1977).

¿Existe un mecanismo biológico posible que vincule la intervención con el evento adverso?

Un evento adverso descrito es más posible si se puede explicar a través de un mecanismo biológico bien comprendido. Por ejemplo, la amiodarona tiene una estructura química similar al yodo, lo que explica los efectos adversos sobre la función tiroidea observados frecuentemente.

¿Los informes proporcionan información suficiente para permitir una evaluación detallada de la evidencia?

Un estudio examinó 1520 informes de casos publicados de supuestas reacciones adversas y encontró diferencias significativas en la información proporcionada en estos informes (Kelly 2003). Con respecto a los detalles de las características de los pacientes, sólo tres variables de los pacientes se describieron más del 90% de las veces, mientras que otras 12 se presentaron menos del 25% de las veces. En la evaluación del fármaco culpable, Kelly encontró que sólo una variable del fármaco (por ejemplo dosis, duración, frecuencia o fórmula exacta) se informó en más del 90% de las veces; otras seis se describieron del 14% al 74% de las veces. La variación significativa en la naturaleza de los informes significa que para los revisores es difícil realizar una evaluación detallada.

¿Existe algún problema potencial con el uso de los datos de los informes que pudiera tener un peso mayor en el beneficio percibido de ser exhaustivos?

Existe un equilibrio entre la conveniencia de “incluir todo” y la necesidad de evitar la publicación de información sesgada o poco fiable que pudiera activar una falsa alarma. El programa de vacunación MMR se interrumpió debido a publicaciones anecdóticas en una revista con buena reputación, y hubo varias personas en el RU dañadas por brotes de sarampión a partir de la disminución de la aplicación de la vacuna (Asaria 2006). La inclusión de información extra (pero potencialmente poco fiable) sobre “eventos adversos” puede tener efectos perjudiciales y los revisores deberán considerar cuidadosamente la repercusión negativa y las ramificaciones legales de agregar esta información.

14.7 Información del capítulo

Autores: Yoon K Loke, Deirdre Price y Andrew Herxheimer del Grupo Cochrane de Métodos de Efectos Adversos.

La versión en inglés de este capítulo se debería citar como: Loke YK, Price D, Herxheimer A. Chapter 14: Adverse effects. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: Los siguientes colegas (enumerados alfabéticamente) han contribuido con su experiencia al Grupo Cochrane de Métodos de Efectos Adversos, y han ayudado a desarrollar esta guía: Jeff Aronson, Anne-Marie Bagnall, Andrea Clarke, Sheena Derry, Anne Eisinga, Su Golder, Tom Jefferson, Harriet MacLehose, Heather McIntosh y Nerys Woolacott.

Recuadro 14.7.a: El Grupo Cochrane de Métodos de Efectos Adversos

El Grupo Cochrane de Métodos de Efectos Adversos (GCMEA) proporciona una guía metodológica sobre las técnicas apropiadas para la identificación y la evaluación sistemática de los efectos adversos. Los orígenes del GCMEA datan de casi una década, de las reuniones informales de algunos individuos involucrados en la evaluación sistemática de los efectos perjudiciales de las intervenciones. Lo anterior llevó, en enero de 2001, a la formación del Subgrupo de Efectos Adversos como parte del Grupo de Métodos de Estudios No Aleatorizados. En junio de 2007 se registró oficialmente el Grupo Cochrane de Métodos de Efectos Adversos (GCMEA).

El principio fundamental del GCMEA es que cada intervención de atención sanitaria conlleva algunos riesgos de daño. Para alcanzar una decisión completamente informada las opciones de tratamiento deberían estar apoyadas por una evaluación sistemática de los beneficios y daños. Las revisiones que se centran principalmente en el beneficio del tratamiento, junto con la falta de información sobre los efectos perjudiciales, crearían dificultades a las personas que tratan de tomar decisiones equilibradas. El GCMEA tiene como objetivo compensar este desequilibrio y colaborar con los Grupos de Revisión y Métodos para mejorar la metodología y la calidad de los análisis de los efectos adversos. El GCMEA estará en disposición de examinar cualquier área de incertidumbre metodológica que requiera investigación adicional, y espera desarrollar y diseminar formas apropiadas de rellenar cualquier laguna que se identifique.

Sitio web: aemg.cochrane.org

14.8 Referencias

Arnaiz 2001
Arnaiz JA, Carne X, Riba N, Codina C, Ribas J, Trilla A. The use of evidence in pharmacovigilance. Case reports as the reference source for drug withdrawals. *European Journal of Clinical Pharmacology* 2001; 57: 89-91.

Asaria 2006
Asaria P, MacMahon E. Measles in the United Kingdom: can we eradicate it by 2010? *BMJ* 2006; 333: 890-895.

Chan 2004
Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004; 291: 2457-2465.

Derry 2001
Derry S, Kong LY, Aronson JK. Incomplete evidence: the inadequacy of databases in tracing published adverse drug reactions in clinical trials. *BMC Medical Research Methodology* 2001; 1: 7.

Edwards 1999
Edwards JE, McQuay HJ, Moore RA, Collins SL. Reporting of adverse effects in clinical trials should be improved: lessons from acute postoperative pain. *Journal of Pain and Symptom Management* 1999; 18: 427-437.

Glasziou 2007
Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007; 334: 349-351.

Glasziou 2004
Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004; 328: 39-41.

Golder 2006
Golder S, McIntosh HM, Duffy S, Glanville J, Centre for Reviews and Dissemination and UK Cochrane Centre Search Filters Design Group. Developing efficient search strategies to identify reports of adverse effects in MEDLINE and EMBASE. *Health Information and Libraries Journal* 2006; 23: 3-12.

Herxheimer 2002
Herxheimer A, Petrie KJ. Melatonin for the prevention and treatment of jet lag. *Cochrane Database of Systematic Reviews* 2002, Issue 2. Art No: CD001520.

Ioannidis 2004
Ioannidis JPA, Evans SJ, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, Moher D. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Annals of Internal Medicine* 2004; 141: 781-788.

Jick 1977
Jick H. The discovery of drug-induced illness. *New England Journal of Medicine* 1977; 296: 481-485.

Jüni 2004
Jüni P, Nartey L, Reichenbach S, Sterchi R, Dieppe PA, Egger M. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *The Lancet* 2004; 364: 2021-2029.

Kelly 2003
Kelly WN. The quality of published adverse drug event reports. *Annals of Pharmacotherapy* 2003; 37: 1774-1778.

Koch-Weser 1977
Koch-Weser J, Sellers EM, Zacest R. The ambiguity of adverse drug reactions. *European Journal of Clinical Pharmacology* 1977; 11: 75-78.

Lanctot 1995
Lanctot KL, Naranjo CA. Comparison of the Bayesian approach and a simple algorithm for assessment of adverse drug events. *Clinical Pharmacology and Therapeutics* 1995; 58: 692-698.

Loke 2004
Loke YK, Derry S, Aronson JK. A comparison of three different sources of data in assessing the frequencies of adverse reactions to amiodarone. *British Journal of Clinical Pharmacology* 2004; 57: 616-621.

Loke 2006
Loke YK, Price D, Derry S, Aronson JK. Case reports of suspected adverse drug reactions--systematic literature survey of follow-up. *BMJ* 2006; 332: 335-339.

McIntosh 2004
McIntosh HM, Woolacott NF, Bagnall AM. Assessing harmful effects in systematic reviews. *BMC Medical Research Methodology* 2004; 4: 19.

Miettinen 1983
Miettinen OS. The need for randomization in the study of intended effects. *Statistics in Medicine* 1983; 2: 267-271.

Olsen 1999
Olsen H, Klemetsrud T, Stokke HP, Tretli S, Westheim A. Adverse drug reactions in current antihypertensive therapy: a general practice survey of 2586 patients in Norway. *Blood Pressure* 1999; 8: 94-101.

Papanikolaou 2006
Papanikolaou PN, Christidi GD, Ioannidis JP. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *Canadian Medical Association Journal* 2006; 174: 635-641.

Schulz 2002
Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. *The Lancet* 2002; 359: 696-700.

Stricker 2004
Stricker BH, Psaty BM. Detection, verification, and quantification of adverse drug reactions. *BMJ* 2004; 329: 44-47.

Vandenbroucke 2004
Vandenbroucke JP. When are observational studies as credible as randomised trials? *The Lancet* 2004; 363: 1728-1731.

Vandenbroucke 2006
Vandenbroucke JP. What is the best evidence for determining harms of medical treatment? *Canadian Medical Association Journal* 2006; 174: 645-646.

Venning 1982
Venning GR. Validity of anecdotal reports of suspected adverse drug reactions: the problem of false alarms. *British Medical Journal (Clinical Research Edition)* 1982; 284: 249-252.

Venning 1983
Venning GR. Identification of adverse reactions to new drugs. II (continued): How were 18 important adverse reactions discovered and with what delays? *British Medical Journal (Clinical Research Edition)* 1983; 286: 365-368.

ÍNDICE
CAPÍTULO 15: INCORPORACIÓN DE PRUEBAS ECONÓMICAS

PUNTOS CLAVE

15.1 LA FUNCIÓN Y RELEVANCIA DE LAS PRUEBAS ECONÓMICAS EN LAS REVISIONES COCHRANE

15.1.1 Introducción
Recuadro 15.1.a: Archie Cochrane y los aspectos económicos de la salud (Cochrane 1972)
15.1.2 Economía y evaluación económica
Recuadro 15.1.b: Tipos de evaluación económica completa
15.1.3 Cobertura de los aspectos económicos en las revisiones Cochrane

15.2 PLANIFICACIÓN DEL COMPONENTE ECONÓMICO DE UNA REVISIÓN COCHRANE

15.2.1 Formulación de una pregunta económica
Figura 15.2.a: Cursos del evento clínico
Recuadro 15.2.a: Comentarios en los antecedentes que destacan los aspectos económicos de las intervenciones
15.2.2 Inclusión de medidas de uso de recursos, costos y costo-efectividad como resultados
15.2.3 Especificación de los tipos de estudios económicos y el alcance del componente económico de una revisión

15.3 LOCALIZACIÓN DE LOS ESTUDIOS

15.3.1 Uso de filtros de búsquedas electrónicas
15.3.2 Uso de bases de datos especializadas

15.4 SELECCIÓN DE LOS ESTUDIOS Y OBTENCIÓN DE LOS DATOS

15.4.1 Evaluación de la relevancia del tema de revisión
15.4.2 Obtención de los datos

15.5 ANÁLISIS DEL RIESGO DE SESGO

15.5.1 Clasificación de los estudios según el diseño de los mismos
15.5.2 Evaluación crítica de la calidad metodológica
Figura 15.5.a: Lista de verificación de Drummond (Drummond 1996)
Figura 15.5.b: Lista de verificación de Evers (Evers 2005)

15.6 ANÁLISIS Y PRESENTACIÓN DE LOS RESULTADOS

Introducció

15.6.1 Presentación de los resultados en tablas
15.6.2 Resumen narrativo de los resultados
15.6.3 Metanálisis de los datos de uso de recursos y costos
Recuadro 15.6.a: Declaración de una decisión de no realizar un metanálisis de datos de uso de recursos o costos

15.6.4 Desarrollo de un modelo económico

15.7 ANÁLISIS DE LOS SESGOS DE NOTIFICACIÓN

15.8 INTERPRETACIÓN DE LOS RESULTADOS

Recuadro 15.8.a: Se destaca la necesidad de estudios económicos adicionales en las conclusiones

15.9 CONCLUSIONES

15.10 INFORMACIÓN DEL CAPÍTULO

Recuadro 15.10.a: El Grupo Campbell y Cochrane de Métodos Económicos

15.11 REFERENCIAS

CAPÍTULO 15: INCORPORACIÓN DE PRUEBAS ECONÓMICAS

Autores: Ian Shemilt, Miranda Mugford, Sarah Byford, Michael Drummond, Eric Eisenstein, Martin Knapp, Jacqueline Mallender, David McDaid, Luke Vale y Damian Walker del Grupo Campbell y Cochrane de Métodos Económicos.

Puntos clave

- La economía es el estudio de la asignación óptima de los recursos limitados para la producción de beneficios para la sociedad, por lo que es relevante para cualquier decisión sanitaria.
- Las decisiones óptimas también requieren las mejores pruebas de efectividad.
- Este capítulo describe los métodos para incorporar las perspectivas y las pruebas económicas a las revisiones Cochrane, con énfasis en la revisión crítica de los estudios económicos en salud.
- La incorporación de las perspectivas y las pruebas económicas a las revisiones Cochrane puede mejorar su utilidad y aplicabilidad para la toma de decisiones en la atención sanitaria y nuevos análisis económicos.

15.1 La función y relevancia de las pruebas económicas en las revisiones Cochrane

15.1.1 Introducción

Las revisiones Cochrane reúnen, seleccionan, critican y combinan datos fiables de múltiples estudios de investigación sobre la efectividad y otros aspectos de las intervenciones de atención sanitaria. Las mismas proporcionan pruebas sólidas sobre la efectividad de las intervenciones, lo que produce información con menos sesgo selectivo y con un mayor poder estadístico, lo cual es más probable que convenza a quienes toman las decisiones, en comparación con los estudios únicos.

Sin embargo, debido a los escasos recursos, a menudo quienes toman las decisiones deben considerar no solamente si una intervención funciona, sino también si su adopción dará a lugar a un uso más eficiente de los recursos. Los temas de las revisiones Cochrane incluyen un amplio rango de preguntas cuyas respuestas son importantes para la mejoría de la salud individual y pública, así como el bienestar en medios en los que los recursos son limitados. Por lo tanto, la cobertura de los aspectos económicos de las intervenciones puede mejorar la utilidad y la aplicabilidad de las revisiones Cochrane como un componente de la base para la toma de decisiones en la atención sanitaria (Lavis 2005).

Durante muchos años se ha argumentado que la promoción de una atención efectiva sin tomar en cuenta los costos de la atención y el valor de cualquier ganancia en la salud puede dar lugar a un uso ineficiente de los fondos públicos y privados asignados a la atención sanitaria, lo cual puede indirectamente provocar daños a los individuos y al público (Williams 1987). De hecho, Archie Cochrane, quien inspiró en gran parte el movimiento de las revisiones sistemáticas (y por supuesto la Colaboración Cochrane) estaba a favor de la toma de decisiones informadas por las pruebas en los aspectos económicos de las intervenciones y por las pruebas sobre su efectividad. El título del trabajo más famoso de Cochrane, su libro de conferencias Rock Carling, es *Effectiveness and Efficiency* (Cochane 1972). El Recuadro 15.1.a contiene dos citas de este libro, lo que ilustra la importancia que Cochrane le dio a la función de las pruebas económicas en la toma de decisiones en la atención sanitaria.

Recuadro 15.1.a: Archie Cochrane y los aspectos económicos de la salud (Cochrane 1972)

> "La asignación de fondos y facilidades se basan casi siempre en las opiniones de expertos, pero cada vez más las necesidades de facilidades adicionales tendrán que basarse en argumentos detallados con "pruebas sólidas" como la ganancia que se espera desde el punto de vista de los pacientes y los costos. Posiblemente haya poco que objetar a lo anterior". (p.82).
>
> "Si alguna vez vamos a obtener resultados "óptimos" de los gastos nacionales del NHS, finalmente debemos ser capaces de expresar los resultados de un tipo de actividad particular en forma de beneficios y costos para la población, y el aumento del beneficio que se obtendría si hubiera más dinero disponible". (p.2).

15.1.2 Economía y evaluación económica

La economía es el estudio de la asignación óptima de los recursos limitados para la producción de beneficios para la sociedad (Samuelson 2005). Los recursos son humanos, de tiempo y habilidades, equipos, instituciones, energía y cualquier otro aporte requerido para implementar y sostener un curso de acción determinado (p.ej. la referencia de un paciente individual a un programa de tratamiento de salud y el tratamiento posterior de las secuelas y complicaciones). Los estudios económicos en salud se definen aquí como estudios de evaluación económica completa, estudios de evaluación económica parcial y estudios únicos de efectividad que incluyen información más limitada con respecto a la descripción, medición o valoración del uso de recursos asociados con las intervenciones.

La evaluación económica completa es el análisis comparativo de los cursos de acción alternativos en términos de costos (uso de recursos) y consecuencias (resultados, efectos) (Drummond 2005). Esta definición distingue la evaluación económica completa de los análisis económicos que se centran solamente en los costos y el uso de recursos, o evaluaciones económicas parciales. La evaluación económica completa no es un método de investigación único; es un marco para estructurar problemas de decisión específicos. Esto significa que el tipo apropiado de evaluación económica completa, y, por lo tanto, el enfoque de la obtención y el análisis de los datos está determinado principalmente por el problema de decisión, o la pregunta económica, según el problema y el punto de vista de quienes toman las decisiones (ver también la Sección 15.2.1). Los estudios de evaluación económica completa tienen como objetivo describir, medir y valorar todos los cursos de acción alternativos relevantes (p.ej. intervención X versus comparador Y), su aporte de recursos y sus consecuencias. Los análisis de costo-beneficio (ACB) se incluyen en esta categoría. Algunos enfoques son valoraciones que no analizan todas las consecuencias pero aún así se consideran evaluaciones económicas completas, incluidos los análisis de costo-efectividad (ACE) y los análisis de costo-utilidad (ACU). Todos los tipos de evaluación económica completa utilizan un enfoque marginal para el análisis. En otras palabras, tienen como objetivo producir medidas de *incremento* del uso de recursos, costos o costo-efectividad. En el Recuadro 15.1.b se proporcionan descripciones breves del ACE, ACU y ACB (ver también el Capítulo 2 de Drummond (Drummond 2005).

Otros tipos de estudios del uso de los recursos de la atención sanitaria no hacen comparaciones explícitas entre las intervenciones alternativas en cuanto a los costos (uso de recursos) y las consecuencias (efectos). Dichos estudios no se consideran evaluaciones económicas completas pero en su lugar se conocen como evaluaciones económicas parciales. Las evaluaciones económicas parciales pueden contribuir con pruebas útiles a la comprensión de los aspectos económicos de las intervenciones. Los estudios económicos en salud que se consideran evaluaciones económicas parciales incluyen análisis de costos, estudios de descripción de costos y descripciones de resultados de costos. Además de las evaluaciones económicas completas y parciales, los ensayos aleatorizados y otros tipos de estudios únicos de efectividad pueden incluir información más limitada con respecto a la descripción, la medición o la valoración del uso de recursos asociados con las intervenciones. Aunque es posible que la inclusión de este tipo de información no siempre constituya un enfoque de evaluación económica completa o parcial, aún puede contribuir con pruebas útiles a la comprensión de los aspectos económicos de las intervenciones.

Los estudios de evaluación económica utilizan y son utilizados en las revisiones sistemáticas de efectos de las intervenciones. Primero, una revisión sistemática puede incluir un componente económico que incorpore una revisión crítica de los estudios económicos en salud publicados y no publicados (ver Sección 15.1.3). Segundo, de la misma manera que un número creciente de evaluaciones económicas completas y parciales se realizan junto con (e incorporan) estudios únicos de efectividad como los ensayos aleatorizados (Maynard 2000, Neumann 2005), las evaluaciones económicas completas también se basan cada vez más en las pruebas de efectos compiladas mediante

métodos de revisiones sistemáticas. De hecho, todos los tipos de evaluaciones económicas completas descritas anteriormente (ACE, ACU, ACB) se pueden realizar junto con, y con la incorporación de, una revisión sistemática de los efectos, incluido el uso de un enfoque de análisis de decisiones para la combinación o el modelado de la evidencia disponible sobre los costos y los efectos de la intervención (Briggs 2006). La evaluación económica se puede considerar en este contexto como una base adicional de síntesis de la evidencia construida sobre el proceso de la revisión sistemática.

Por lo tanto, las revisiones Cochrane y otras revisiones sistemáticas pueden proporcionar una fuente de datos útil para describir los ejercicios de modelado de la evaluación económica completa posterior o paralela, independientemente de si la revisión incorpora o no una cobertura adicional de los aspectos económicos de las intervenciones. En particular, se ha propuesto que un metanálisis de los datos bien realizado sobre el tamaño del efecto, los efectos adversos y las complicaciones, a partir de una revisión sistemática de ensayos aleatorizados, es la fuente menos sesgada de datos para proporcionar el tamaño del efecto y los parámetros de efectos adversos en un modelo económico (Cooper 2005). Esto se debería complementar con búsquedas sistemáticas adicionales en fuentes de datos apropiadas para describir los rangos de valores para los otros parámetros clave en la fórmula de costo-efectividad o el modelo económico (Weinstein 2003, Philips 2004, Cooper 20059.

Recuadro 15.1.b: Tipos de evaluación económica completa

Todos los tipos de evaluación económica completa comparan los costos (uso de recursos) asociados con una o más intervenciones alternativas (p.ej. intervención X versus comparador Y) con sus consecuencias (resultados, efectos). Todos los tipos valoran los recursos de la misma manera (es decir, al aplicar los costos unitarios a las unidades de uso de recursos medidas). Los tipos difieren principalmente en la manera en la que ellos caracterizan y valoran los efectos. Estas diferencias reflejan los diversos objetivos y puntos de vista de los diferentes problemas de decisión (o preguntas económicas).

Análisis de costo-efectividad (ACE): los efectos de una intervención (y sus comparadores) se miden en unidades de resultado idénticas (p.ej. mortalidad, infartos de miocardio, función pulmonar, peso, hemorragia, infecciones secundarias, cirugías para revisión). Las intervenciones alternativas se comparan según el “costo por unidad de efecto”.

Análisis de costo-utilidad (ACU): cuando las intervenciones alternativas producen diferentes niveles de efectos en términos de cantidad y calidad de vida (o efectos diferentes), los efectos se pueden expresar en utilidades. Las utilidades son medidas que incluyen la duración de la vida y los niveles subjetivos de bienestar. La medida de utilidad que se conoce mejor es los años de vida ajustados por la calidad, o QALY (por sus siglas en inglés). Las intervenciones alternativas se comparan en términos de costo por unidad de utilidad ganada (p.ej. costo por QALY).

Análisis de costo-beneficio (ACB): cuando los aportes de recursos y los efectos de las intervenciones alternativas se expresan en unidades monetarias, de manera que se comparan directamente y entre los programas dentro del sistema de atención sanitaria, o con programas fuera de la atención sanitaria (p.ej. intervención de atención sanitaria versus intervención en la justicia criminal).

15.1.3 Cobertura de los aspectos económicos en las revisiones Cochrane

El objetivo general de este capítulo es describir cómo los autores de las revisiones Cochrane y otras revisiones sistemáticas pudieran compilar la mejor evidencia sobre los aspectos económicos de las intervenciones además de la mejor evidencia de su efectividad.

Actualmente no hay una exigencia formal para que las revisiones Cochrane incluyan los aspectos económicos. Por lo tanto, estas directrices se presentan como una serie de métodos opcionales a considerar por los revisores Cochrane que planifican incluir aspectos económicos. El elemento principal del marco metodológico esbozado es una revisión crítica de los estudios económicos en salud, que se puede realizar como un componente completamente integrado de una revisión Cochrane. Esto incluye la obtención, selección, evaluación crítica, el resumen y la posible síntesis de los datos de estudios económicos en salud relevantes. Las tres premisas fundamentales de estas directrices son las siguientes:

1. Debido a que los usuarios finales de las revisiones Cochrane forman un auditorio internacional, el objetivo general de los componentes económicos de las revisiones debe ser resumir lo que se conoce de diferentes ámbitos acerca de los aspectos económicos de las intervenciones, para ayudar a los usuarios finales a comprender las compensaciones económicas clave entre pruebas o tratamientos sanitarios alternativos;

2. Los objetivos secundarios clave son proporcionar un marco para que las revisiones Cochrane presenten datos económicos y clínicos en un formato que facilite su uso en análisis económicos posteriores o paralelos;

3. Los aspectos económicos son relevantes para la toma de decisiones incluso cuando las pruebas de efectividad de la intervención son inciertas. Primero, los usuarios finales a menudo necesitan conocer las pruebas con respecto al aumento del uso de recursos y de los costos asociados con una intervención versus comparadores relevantes, ya que esto puede ayudar a aclarar dónde invertir en investigaciones futuras sobre efectividad y costo-efectividad. Segundo, es importante para los usuarios finales conocer si las evaluaciones económicas completas existentes se basan o no en pruebas sólidas con respecto a su efectividad.

Los autores de las revisiones Cochrane que planifican incluir aspectos económicos de las intervenciones necesitarán considerar en detalle y desde los estadios más tempranos de desarrollo del protocolo, cómo se relacionan los aspectos económicos con su tema de revisión específico. El uso de los métodos descritos en este capítulo también requerirá de al menos cierto entrenamiento en el uso de métodos económicos en la salud. Por lo tanto, una vez que se toma la decisión de incluir aspectos económicos, es recomendable consultar lo antes posible a un economista en salud con experiencia en los métodos de las revisiones sistemáticas.

Algunos Grupos Cochrane de Revisión (GCR) ya tienen acceso a uno o más economista de salud experimentado que contribuye regularmente con el trabajo de los componentes económicos de las revisiones. El Grupo Campbell y Cochrane de Métodos Económicos (GCCME) ayudará a los autores de revisiones Cochrane a identificar economistas en salud que contribuyan con este trabajo, o proporcionará asesoría o apoyo de revisores pares (ver Recuadro 15.10.a).

15.2 Planificación del componente económico de una revisión Cochrane

15.2.1 Formulación de una pregunta económica

Después de tomar la decisión de incluir aspectos económicos de las intervenciones en una revisión Cochrane, el primer estadio de la investigación es formular una o más preguntas, u objetivos, que se intenten abordar mediante el componente económico de la revisión. Cada pregunta u objetivo económico determinará las decisiones metodológicas en los estados posteriores de la revisión crítica de los estudios económicos en salud.

La formulación de una pregunta económica requiere una cuidadosa consideración de la función y la relevancia de los aspectos económicos para el tema específico de la revisión. Las siguientes preguntas *preliminares* pretenden proporcionar puntos de partida útiles que ayuden a los revisores y editores a conceptualizar la función y la relevancia de los aspectos económicos.

- ¿Cuál es la carga económica para la sociedad (p.ej. sistema de salud, proveedores de atención sanitaria o social, individuos, familias, empleados) de la afección o enfermedad sobre la cual la intervención intenta influir?

- ¿Qué tipos de aumentos del aporte de recursos se necesitan para implementar y sostener la intervención en comparación con los comparadores (p.ej. personal, equipos, fármacos, atención intrahospitalaria)?

- ¿Cuáles son las consecuencias del aumento del aporte de recursos al implementar la intervención en comparación con los comparadores? o ¿Cómo pudiera la intervención repercutir sobre el uso posterior (consiguiente) de los recursos en comparación con los comparadores (p.ej. complicaciones, procedimientos secundarios, visitas ambulatorias, tiempo fuera del trabajo)?

- ¿Cuáles son los incrementos en los costos asociados con los cambios en el uso de recursos que pueden ser consecuencia de la intervención en comparación con los comparadores (p.ej. costos médicos directos e indirectos, gastos realizados por el paciente, ingresos de empleo)?

- ¿Cuál es el valor económico asociado con el aumento del beneficio o los efectos (resultados) adversos que pueden ser consecuencia de la intervención en comparación con los comparadores (p.ej. medidas de voluntad para pagar, o utilidad)?

- ¿Cuál es el posible balance entre los costos (uso de recursos) y los efectos beneficiosos o adversos que puede ser necesario considerar en una decisión para adoptar o rechazar un curso de acción determinado?

Al considerar estas preguntas preliminares, es importante tomar en cuenta los siguientes aspectos clave:

- Magnitud: ¿Cuál es la probable magnitud de los diferentes ítems del aumento del uso de recursos o el aumento de los costos asociados con la intervención en comparación con los comparadores? En otras palabras, ¿qué ítems del uso de los recursos (aporte de recursos y consecuencias de los recursos) y qué costos es probable que sean los más importantes cuando se toman decisiones entre intervenciones alternativas?

- Horizonte de tiempo: ¿Cuál es el horizonte de tiempo en el cual es probable que se acumulen los costos (uso de recursos) y los efectos (resultados) importantes? Las revisiones Cochrane establecen de manera implícita un horizonte de tiempo para los efectos al especificar medidas intermedias y finales de evaluación de los efectos como

medidas de resultado proyectadas. Existe una necesidad paralela de considerar si el mismo horizonte de tiempo es aplicable cuando se consideran juntos todos los costos (uso de recursos) y efectos relevantes.

- Punto de vista analítico: ¿Quién es probable que asuma el aumento de los costos asociados con la intervención en comparación con los comparadores, y quién recibe el aumento de los beneficios (p.ej. paciente, familia del paciente, proveedor de atención sanitaria o un tercero, el sistema sanitario, la sociedad)? Algunos costos (uso de recursos) son relevantes desde un punto de vista analítico, pero no desde otro. Por ejemplo, el costo de proporcionar atención informal puede ser relevante desde el punto de vista de un paciente o la sociedad, pero es posible que se excluya cuando se selecciona una perspectiva más estrecha, como la del sistema de atención sanitaria. Una complicación adicional es que cierto uso de los recursos o ciertas categorías de costos se pueden superponer entre las perspectivas. Debido a la variedad de usuarios finales de las revisiones Cochrane, un enfoque pragmático es considerar el rango completo de perspectivas y luego notificar no solamente las medidas de uso de recursos y costos, sino también quién asume el costo o quién utiliza los recursos.

El curso de un evento clínico puede proporcionar una herramienta adicional útil que ayude a conceptualizar la función y la relevancia de los aspectos económicos para un tema específico de revisión. El curso de un evento clínico proporciona un método sistemático y explícito de representar diferentes procesos y resultados de la atención sanitaria y social. El método incluye la descripción de los cursos principales de los eventos que tienen diferentes implicaciones de recursos o valores de resultados asociados con ellos, desde el momento de la introducción de las intervenciones, a través de cambios posteriores en el tratamiento de los participantes, hasta los resultados finales (ver también Capítulo 2 de Donaldson (Donaldson 2002). La Figura 15.2.a muestra un ejemplo del curso de un evento clínico para el “accidente cerebrovascular”. Al desarrollar un curso de un evento clínico, nuevamente es importante considerar los aspectos clave de la magnitud, el horizonte de tiempo y el punto de vista analítico.

Una vez que se han considerado cuidadosamente la función y la relevancia de los aspectos económicos, es posible formular una o más preguntas u objetivos económicos. Los revisores deben evitar formular preguntas económicas como “¿Cuál es el costo-efectividad de la intervención X (comparada con Y o Z)?”, ya que es poco probable que una revisión crítica de los estudios económicos de la salud proporcione una respuesta fiable a este tipo de pregunta que sea aplicable entre los ámbitos. Las preguntas u objetivos económicos se deben plantear de forma explícita en la sección “Objetivos” del protocolo de una revisión, junto con otras preguntas y objetivos de investigación.

Las consideraciones de la función y la relevancia de los aspectos económicos también se pueden utilizar para los comentarios sobre los aspectos económicos de las intervenciones, e incluirlos en la sección “Antecedentes” de la revisión.

Es posible incluir un “comentario económico” independientemente de si los revisores pretenden o no incorporar una revisión crítica de estudios económicos en salud. Esto es útil para ayudar a colocar a las intervenciones en estudio en un contexto económico y destacar sus posibles consecuencias económicas para la consideración de los usuarios finales de la revisión. Los “comentarios económicos” pueden destacar la carga económica de la enfermedad o la afección médica abordada por las intervenciones, los tipos de recursos necesarios para implementar y sostener las intervenciones (aporte de recursos), las repercusiones potenciales de las intervenciones en el uso posterior de los recursos (consecuencias de los recursos) y aspectos de costo-efectividad. El comentario debe estar apoyado por referencias apropiadas y comentarios críticos sobre la bibliografía relevante, de ser posible. El Recuadro 15.2.a muestra algunos ejemplos de este tipo de comentarios, extraídos de las secciones “Antecedentes” de revisiones Cochrane actuales.

Figura 15.2.a: Cursos del evento clínico

Cursos del evento	Ejemplo
Evento clínico.	Accidente cerebrovascular.
↓	↓
Tratamiento del evento clínico + eventos clínicos posteriores.	Atención aguda y rehabilitación + secuelas y complicaciones del tratamiento.
↓	↓
Recursos utilizados para tratar los eventos y los resultados de los eventos.	Duración de la estancia hospitalaria, intensidad de la terapia de rehabilitación, tratamiento de las secuelas y complicaciones (p.ej. hemorragia a partir de la profilaxis secundaria) y resultados de salud asociados con cada estadio.
↓	↓
Costo de los recursos utilizados y utilidades de los resultados.	Valoración de los recursos mediante el pago (y otros) y los precios de la atención sanitaria y valoración de los resultados, por ejemplo, mediante los años de vida ajustados por la calidad o la voluntad de pagar.

Recuadro 15.2.a: Comentarios en los antecedentes que destacan los aspectos económicos de las intervenciones

"La incontinencia fecal… puede ser un problema debilitante con implicaciones médicas, sociales y económicas… En los Estados Unidos se gastan anualmente más de $400 millones en numerosos productos para la incontinencia fecal y urinaria … Durante 1991 los costos directos de los apósitos, accesorios y otros ítems prescritos en los hospitales y los centros sanitarios de larga estancia en el RU para la incontinencia en general se calculó en £68 millones... Con el aumento del número de personas de edad avanzada en el mundo, esta afección será un reto creciente para los servicios de atención sanitaria y los cuidadores en el domicilio." (Brown 2007).

"Si este nuevo y relativamente costoso tratamiento (Lamotrigina) va a estar disponible para su uso habitual, es necesaria una clara comprensión de cómo se compara con un fármaco antiepiléptico estándar (FAE) como la carbamazepina. Las posibles implicaciones de costos se destacan en una encuesta de los servicios de epilepsia en el North West, RU, que mostró que casi el 40% de los costos de los fármacos (el mayor contribuyente único a los costos directos de la epilepsia) se le atribuyó a los FAE modernos lamotrigina y vigabatrina, a pesar del hecho de que sólo los tomaba el 7% de los pacientes." (Gamble 20069.

"El costo del tratamiento de quimioterapia paliativa para el cáncer colorrectal avanzado incluye no sólo los costos asociados con la administración de quimioterapia, sino también la provisión de apoyo para tratar las complicaciones relacionadas con la quimioterapia. Si la quimioterapia paliativa mejora el control de los síntomas y la calidad de vida, se puede reducir la dependencia del paciente y la necesidad de otras medidas de atención sintomáticas/de apoyo que compensen el costo de este tratamiento. Por otra parte, la incidencia de toxicidad relacionada con la quimioterapia es alta, y hay una disminución en la calidad de vida debido al tratamiento, luego los costos de la quimioterapia paliativa serán mucho mayores que los de la atención de apoyo sola." (Best 2000).

15.2.2 Inclusión de medidas de uso de recursos, costos y costo-efectividad como resultados

El proceso de formulación de las preguntas económicas también puede ayudar a aclarar el grupo de medidas importantes de uso de recursos, costos o costo-efectividad (o una combinación de ellas) a incluir como resultados proyectados en una revisión. Estos resultados se deben incluir junto con otros resultados proyectados en el apartado "Tipos de medidas de resultado" de la sección de la revisión "Criterios para considerar los estudios para esta revisión". Cuando sea posible, es útil separar las medidas de uso de recursos y costos al nivel de ítems o categorías específicas (p.ej. duración de la estancia hospitalaria en días, duración de la operación en minutos, número de consultas ambulatorias, hemorragias debido a la profilaxis secundaria a los seis meses de seguimiento, número de días fuera del trabajo, uso directo de recursos médicos, costos médicos directos, uso indirecto de recursos médicos o costos, gastos para el paciente) y evitar el uso de términos descriptivos generales para los resultados (p.ej. "costos", "utilización de recursos", "aspectos económicos sanitarios"). Entre las medidas de costo-efectividad que se pueden incluir como medidas de resultado proyectadas en una revisión se incluyen el aumento de las tasas de costo-efectividad, el aumento del costo por los años de vida ajustados por la calidad y las tasas de costo-beneficio (ver también Sección 15.1.2).

15.2.3 Especificación de los tipos de estudios económicos y el alcance del componente económico de una revisión

Una revisión crítica de los estudios económicos en salud debe especificar al inicio qué tipo de estudios se considerarán para inclusión (ver también Sección 15.1.2). Esta decisión está regida principalmente por las preguntas u objetivos económicos que se han formulado y las medidas de uso de recursos, costos y costo-efectividad incluidas como medidas de resultado seleccionadas.

Esta decisión se debe tomar en consulta con un economista en salud, ya que no es necesariamente un ejercicio sencillo trazar el mapa del curso analítico entre diferentes formas de preguntas económicas, medidas de resultado "económicas" y diferentes tipos de estudios económicos en salud. Por ejemplo, si un análisis de costo-efectividad incluye los resultados de todos los estadios intermedios del análisis junto con los resultados finales, es posible extraer los datos de resultado relacionados con las medidas de uso de recursos, costos y costo-efectividad; sin embargo, si sólo se describen los resultados finales, puede que sólo sea posible extraer los datos de resultado relacionados con las medidas de costo-efectividad.

Los tipos de estudios económicos en salud a considerar para su inclusión en la revisión se deben señalar en el apartado "Tipos de estudios" de la sección "Criterios para considerar los estudios en esta revisión". Un planteamiento ilustrativo que muestra la variedad completa de estudios económicos es el siguiente:

> Tipos de estudios
> Los siguientes tipos de estudios se considerarán para inclusión en la revisión crítica de los estudios económicos en salud:
> Estudios de evaluación económica completa (es decir, análisis de costo-efectividad, análisis de costo-utilidad, análisis de costo-beneficio) de (intervención/es versus comparador/es); evaluaciones económicas parciales (es decir, análisis de costos, estudios de descripción de costos, descripciones de resultados de costos) de (intervención/es y comparador/es); y ensayos aleatorizados que proporcionen

información más limitada, como las estimaciones de uso de recursos o costos asociados con (intervención/es y comparador/es).

Una decisión metodológica clave final cuando se planifica una revisión crítica de estudios económicos en salud es establecer el alcance de este elemento en el proceso de revisión. Hay al menos tres opciones para el alcance de una revisión crítica de estudios económicos en salud:

1. Considerar solamente los estudios económicos en salud relevantes realizados junto con estudios de efectividad que cumplan los criterios de elegibilidad para el componente de efectividad de la revisión;

2. Considerar los estudios económicos en salud realizados junto con estudios de efectividad (también basados en datos que provienen de éstos) que cumplen los criterios de elegibilidad para el componente de efectividad de la revisión;

3. Considerar los estudios económicos en salud, independientemente de si se realizaron junto con estudios de efectividad (o basados en éstos) que cumplen los criterios de elegibilidad para el componente de efectividad de la revisión.

Generalmente la primera opción pudiera permitir solamente estudios económicos en salud realizados junto con ensayos aleatorizados de alta calidad que se considerarán para inclusión en el componente económico de la revisión. La segunda opción permitiría *adicionalmente* la consideración de estudios con modelos económicos basados en un metanálisis de datos de ensayos aleatorizados de alta calidad. Un buen ejemplo de una revisión de modelos económicos es la revisión del cribado para el aneurisma de la aorta abdominal realizada por Campbell y colegas (Campbell 2007). La tercera opción es claramente más inclusiva y permite la consideración de todos los estudios económicos en salud, incluidos los que se basan en estudios observacionales de bases de datos administrativas grandes, o análisis de costos basados en regresión y uso de recursos, por ejemplo.

Se conoce poco acerca de la repercusión de la inclusión de estos tipos diferentes de estudios económicos en salud sobre los resultados de una revisión crítica. Sin embargo, es posible que este tipo de decisión con respecto al “alcance” tenga al menos la posibilidad de repercutir en los resultados, ya que diferentes opciones pueden incluir la consideración de diferentes grupos de estudios (ver también Sección 15.5.2). Además, cuando la revisión incluye evaluaciones económicas basadas en estudios únicos (p.ej. ensayos aleatorizados) y evaluaciones económicas basadas en modelos, puede ser óptimo considerar cada una de estas categorías de estudios por separado, con el fin de mantener la comparabilidad entre los estudios.

En la práctica, la mayoría de las revisiones Cochrane actuales que deciden incorporar la inclusión de evidencia de estudios económicos en salud limitan esta inclusión a los estudios económicos realizados junto con los estudios de efectividad que cumplen los criterios de elegibilidad para el componente de efectividad de la revisión (es decir, la primera opción), pero no lo señalan explícitamente (Shemilt 2007). Debido a que es posible que la decisión con respecto al alcance excluya algunos estudios económicos en salud en los que no sea posible la evaluación crítica de su calidad metodológica, el resultado de esta decisión se debe señalar en el apartado “Tipo de estudios” de la sección de la revisión “Criterios para considerar los estudios para esta revisión”. Juntamente se describirán los detalles de los tipos de estudios económicos que se considerarán para inclusión, por ejemplo agregando “La revisión considerará solamente los estudios económicos realizados junto con estudios de efectividad incluidos en el componente de efectividad de la revisión” al planteamiento anterior.

15.3 Localización de los estudios

15.3.1 Uso de filtros de búsquedas electrónicas

Los métodos de búsqueda para localizar estudios económicos en salud relevantes diferirán según el alcance de una revisión crítica de dichos estudios y de los tipos de estudios que se considerarán para inclusión (ver también Secciones 15.2.3 y 15.1.2). Sin embargo, en todos los casos el primer estadio de la estrategia de búsqueda tendrá el mismo objetivo: identificar los estudios de efectividad recuperados para un examen inicial y su inclusión potencial en una revisión Cochrane que incluya estudios económicos en salud relevantes.

Los registros electrónicos de estudios de efectividad recuperados de las bases de datos bibliográficas electrónicas se pueden filtrar mediante estrategias de búsqueda diseñadas para obtener los estudios económicos. Esto puede preceder al examen visual de los resúmenes y los textos completos de los estudios y funcionar como una ayuda para localizar los estudios económicos al limitar el número de registros a evaluar. El filtrado electrónico es más útil en revisiones en las que el número de registros recuperados de las bases de datos electrónicas es grande (es decir, en los que el número es relativamente pequeño, es posible que el uso de filtros electrónicos no se considere necesario, pero aún será necesario aplicar criterios explícitos).

El *Centre for Reviews and Dissemination* (CRD) ha desarrollado una serie de estrategias de búsquedas electrónicas para obtener los posibles estudios de evaluación económica para su inclusión en la *NHS Economic Evaluation Database* (NHS EED). Las versiones de MEDLINE (Ovid CD-ROM), CINAHL (Ovid CD-ROM), EMBASE (Ovid en línea) y PsychINFO (Ovid en línea) se publican en el NHS EED Handbook (Craig 2007) y en línea en www.york.ac.uk/inst/crd/nfaq2.htm. Cada una de estas estrategias de búsqueda se puede anexar a las estrategias de búsqueda específicas de la base de datos correspondiente mediante el operador “AND”, para filtrar los resultados de la búsqueda para los resultados que también contengan términos de búsqueda “económicos”.

Estas estrategias de búsqueda NHS EED son muy amplias y obtendrán estudios de métodos económicos y revisiones de estudios económicos, así como un rango completo de tipos de estudios económicos en salud (ver Sección 15.1.2). Para búsquedas más específicas se recomiendan adaptaciones más estrechas de las estrategias de búsqueda y una lectura cuidadosa de las notas de alcance de los términos MeSH. Las estrategias de búsqueda también se pueden adaptar, después de consultar con especialistas en recuperación de información, para su uso en otras bases de datos bibliográficas electrónicas. La adaptación de las estrategias de búsqueda deberá tomar en cuenta las variaciones entre las bases de datos en la indexación o clasificación de los estudios económicos en salud. Está disponible una lista comentada útil de bases de datos bibliográficas electrónicas que incluye la bibliografía económica en salud y detalles de sitios de Internet que contienen bibliografía gris relevante (Napper 2005).

Una consideración importante con respecto al procedimiento cuando se considera el uso de filtros de búsqueda electrónica diseñados para obtener estudios económicos en salud es que las revisiones Cochrane también utilizan frecuentemente otros filtros de búsqueda diseñados para obtener otros diseños de estudios específicos, como los ensayos aleatorizados. Estos “filtros de búsqueda de diseños de estudios” también se anexan a las estrategias de búsqueda específicas para la revisión mediante un operador “AND”. Por lo tanto, si el alcance de la revisión crítica no se limita a los estudios económicos en salud realizados junto con estudios de efectividad incluidos en el componente de efectividad de la revisión (p.ej. evaluaciones económicas basadas en modelos; ver Sección 15.2.3), el “filtro de búsqueda económica” se debe anexar a cualquier otro “filtro de búsqueda de diseños de estudios” mediante el operador “OR” para asegurar que se consideren y obtengan todos los tipos de estudios económicos en salud. Alternativamente, si el alcance

de la revisión crítica se limita a los estudios económicos en salud realizados junto con los estudios de efectividad incluidos en el componente de efectividad de la revisión, no es necesario el uso de "filtro de búsqueda económica", ya que la mayoría de los estudios económicos a considerar se recuperarán mediante el "filtro de búsqueda de diseños de estudios" (no obstante, es posible que en este caso los resultados de la búsqueda aún puedan omitir algunos estudios económicos relevantes, como las evaluaciones económicas basadas en ensayos aleatorizados pero publicadas por separado y habitualmente después de los resultados del ensayo).

15.3.2 Uso de bases de datos especializadas

La *NHS Economic Evaluation Database* (NHS EED) se publica como parte de *The Cochrane Library* (www.thecochranelibrary.com). Por lo tanto, siempre que los usuarios realicen búsquedas en *The Cochrane Library*, se señalarán los registros NHS EED, así como las revisiones Cochrane. La NHS EED también está disponible gratis en línea en el sitio web del Centre for Reviews and Dissemination (CRD) (ver www.york.ac.uk/inst/crd/crddatabases.htm). La versión de la NHS EED en *The Cochrane Library* se actualiza trimestralmente, mientras que la versión del sitio web del CRD se actualiza mensualmente.

Se recomienda una búsqueda en la NHS EED y el procesamiento de los resultados de esta búsqueda para todas las revisiones Cochrane, especialmente para las que incorporen una revisión crítica de estudios económicos en salud. La NHS EED contiene resúmenes estructurados de evaluaciones económicas completas en la atención sanitaria, publicadas en cualquier idioma, así como los registros bibliográficos de evaluaciones económicas parciales, estudios metodológicos y revisiones de estudios económicos. El formato de los resúmenes estructurados de la NHS EED incluye un comentario crítico redactado por un revisor independiente economista en salud, y presenta detalles de los métodos, resultados y otros datos en un formato resumen que es útil para redactar la evaluación crítica y la obtención de los datos en una revisión crítica de estudios económicos en salud (ver Secciones 15.5.2 y 15.4.2).

En ocasiones se puede considerar útil incluir los resúmenes NHS EED de estudios de evaluación económica completa relevantes como un anexo de una revisión Cochrane publicada, como lo hicieron Rodgers y cols. y Fayter y cols. (Rodgers 2006, Fayter 2007) (ver también Sección 15.6.2). Si la NHS EED no contiene un resumen estructurado de una evaluación económica completa identificada durante las búsquedas realizadas para una revisión Cochrane, sería útil que los revisores pudieran alertar al Grupo Campbell y Cochrane de Métodos Económicos (Recuadro 15.10.a), de manera que quienes busquen en NHS EED puedan estar al tanto de la necesidad de considerar la producción de un resumen.
Las búsquedas en NHS EED y otras bases de datos especializadas de bibliografía económica en salud (ver más adelante) se pueden realizar mediante adaptaciones de estrategias de búsqueda específicas para la revisión, en las que se excluyen los "filtros de búsqueda económica" y otros "filtros de búsqueda de diseños de estudio". Cuando se busca en The Cochrane Library, se realizan búsquedas por defecto en NHS EED (es decir, a menos que la base de datos se excluya específicamente de la búsqueda mediante las opciones de búsqueda avanzada). La información sobre cómo buscar en la versión del sitio web del CRD del NHS EED se puede obtener en las páginas de ayuda del CRD en www.crd.york.ac.uk/crdweb/html/help.htm.

El deseo de extender los principios de la base de datos NHS EED establecida en el RU a otros países europeos ha dado lugar al establecimiento de las *European Network of Health Economic Evaluation Databases* (EURONHEED), que también están disponible gratuitamente en línea (ver http://infodoc.inserm.fr/euronheed/). La NHS EED proporciona vínculos solamente a registros de resúmenes completos de EURONHEED (desde 2000 en

adelante), de manera que aunque las búsquedas en NHS EED recuperarán todos los registros de resúmenes completos de ambas bases de datos, no recuperarán los registros bibliográficos de evaluaciones económicas parciales, estudios metodológicos o revisiones de estudios económicos que estén solamente en EURONHEED.

NHS EED, EURONHEED y otras bases de datos especializadas de bibliografía económica en salud en las que se puede buscar para las revisiones Cochrane (incluidas *The CEA Registry, Health Economic Evaluations Database* (HEED) y *Econlit*) se describen completamente en un artículo publicado por el equipo de proyecto de NHS EED (Aguiar-Ibañez 2005). El CRD también publicó una lista comentada en línea que contiene detalles de estas bases de datos, incluidos los vínculos a cada sitio web de la base de datos, en www.york.ac.uk/inst/crd/econ4.htm, como parte de sus páginas "Recursos de información en economía en salud" (www.york.ac.uk/inst/crd/econ.htm). Esta lista comentada también incluye detalles de las bases de datos generales seleccionadas en las que se encuentra la bibliografía económica en salud (ver también Sección 15.3.1).

Si el alcance de la revisión crítica de los estudios económicos en salud se limita a estos estudios realizados junto con los estudios de efectividad que cumplen los criterios de elegibilidad para el componente de efectividad de la revisión (ver Sección 15.2.3), entonces el único objetivo de una búsqueda complementaria en NHS EED y otras bases de datos especializadas es verificar si incluyen cualquier resumen estructurado de estudios de evaluación económica completa realizados junto con los estudios de efectividad incluidos. Sin embargo, si el alcance de la revisión crítica de estudios económicos en salud es más amplio (ver Sección 15.2.3), entonces el objetivo *adicional* es identificar estudios económicos *adicionales* para su posible inclusión en la revisión.

15.4 Selección de los estudios y obtención de los datos

15.4.1 Evaluación de la relevancia del tema de revisión

Una vez que se han obtenido los artículos de texto completo de los estudios económicos en salud potencialmente relevantes (y los resúmenes estructurados de las evaluaciones económicas completas, cuando estén disponibles), el siguiente paso es evaluar la relevancia de cada uno de estos estudios para el tema específico de la revisión, como un estadio preliminar para abordar el aspecto del riesgo de sesgo. Las decisiones de incluir o excluir estudios económicos en salud según la relevancia se deben basar en si cumplen o no los criterios de elegibilidad con respecto a las poblaciones objetivo, las intervenciones, las comparaciones y los resultados que se especificaron en el protocolo de la revisión. Los motivos para excluir estudios económicos en salud en este estadio se deben informar en las tablas "Características de los estudios excluidos".

15.4.2 Obtención de los datos

Es necesario especificar para cada revisión individual los requisitos para la obtención precisa de los datos para los componentes económicos de las revisiones Cochrane, según la pregunta u objetivo económico específico y las medidas de *aumento* del uso de recursos, costos o costo-efectividad incluidas como resultados proyectados. En general, será necesario obtener dos tipos de datos: los detalles de las características de los estudios económicos en salud incluidos y los detalles de sus resultados. La posibilidad de obtener los datos de las publicaciones, como se indica más delante, puede estar limitada por la calidad de los estudios económicos en salud (cuando hay información faltante, una opción adicional es establecer contacto con los autores de los estudios para solicitar detalles adicionales).

Es probable que entre los datos útiles a obtener con respecto a las características de cada estudio económico se incluyan: año del estudio; detalles de las intervenciones y comparadores; diseño del estudio y fuente/s de uso de recursos, costos por unidad y (de ser aplicable) datos de efectividad (ver también Secciones 15.1.2 y 15.2.3); ámbito jurisdiccional, geográfico y organizacional de la toma de decisiones; punto de vista analítico; y horizonte de tiempo para los costos y los efectos (ver Sección 15.2.1).

Para los resultados, las estimaciones de los ítems específicos del uso de recursos asociado con las intervenciones y los comparadores y las estimaciones del costo por unidad se deben extraer de forma separada, si se describen, así como las estimaciones de los costos del uso de recursos (es decir, el número de unidades de recurso X costo por unidad). El tipo y cantidad de cada recurso utilizado se debe extraer en unidades naturales (p.ej. duración de la estancia hospitalaria en días, duración de la operación en minutos, número de visitas ambulatorias a los seis meses de seguimiento, número de días de trabajo). También es importante obtener información sobre el precio del año y la moneda utilizada para calcular las estimaciones de los costos y el aumento de estos. Las medidas de aumento del uso de recursos y los costos se deben obtener a nivel del paciente individual (es decir, uso de recursos por paciente, costo por paciente), siempre que sea posible. Se debe extraer una estimación puntual y una medida de incertidumbre (p.ej. error estándar o intervalo de confianza) para las medidas de aumento del uso de recursos, costos y costo-efectividad, si se detallan. Adicionalmente, es útil obtener detalles de cualquier análisis de sensibilidad realizado y de cualquier información con respecto a la repercusión de las diferentes suposiciones sobre la magnitud y la dirección de los resultados.

El CRD Report 6 (Craig 2007) incluye una plantilla para producir resúmenes estructurados de evaluaciones económicas completas para su inclusión en NHS EED (ver también Sección 15.3.2), junto con notas para guiar la obtención de los datos y la evaluación crítica. Estos materiales pueden proporcionar una plantilla útil para el diseño de los formularios de obtención de datos para su uso en los componentes económicos de las revisiones Cochrane.

Si una evaluación económica completa ya tiene un resumen estructurado NHS EED correspondiente, se *puede* obviar la necesidad de que los investigadores realicen una obtención adicional de datos del estudio. Al mismo tiempo, debido a que para una revisión Cochrane será necesario realizar la evaluación crítica y la obtención de datos de los estudios de evaluación económica sin un resumen NHS EED completo, se les recomienda a los revisores que consideren registrarse en NHS EED para producir un resumen, con el fin de evitar la duplicación de esfuerzos. Pueden establecer contacto con el GCCME para información adicional, o iniciar la solicitud de que el MHS EED produzca un resumen estructurado (ver también Sección 15.3.2).

15.5 Análisis del riesgo de sesgo

15.5.1 Clasificación de los estudios según el diseño de los mismos

Como estadio preliminar al análisis del riesgo de sesgo se deben clasificar los estudios económicos en salud según el diseño de los mismos. Los métodos que fundamentan la evaluación crítica de la calidad metodológica de los estudios económicos en salud variarán ligeramente según el diseño del estudio.

La clasificación debe consistir de dos estadios:
1. Clasificación del diseño del estudio económico en salud.

2. Clasificación del diseño del estudio que genera los datos de efectividad sobre los cuales se basan los estudios económicos en salud, de ser aplicable.

Cada estudio económico en salud se puede clasificar (estadio 1) como un tipo de evaluación económica completa, un tipo de evaluación económica parcial o un tipo de estudio de efectividad (p.ej. un ensayo aleatorizado) que proporciona información más limitada del uso de recursos o los costos asociados con una intervención (ver Sección 15.1.2). La clasificación del diseño del estudio que genera los datos de efectividad en los cuales se basa el estudio económico en salud (estadio 2) sólo es aplicable en el caso de los estudios económicos en salud clasificados como una evaluación económica completa o una descripción del costo-resultado en el primer estadio de la clasificación. El estudio que genera los datos de efectividad puede tener un diseño de estudio único (p.ej. ensayo aleatorizado, ensayo no aleatorizado, estudio observacional) o una síntesis de varios estudios (p.ej. un metanálisis de ensayos aleatorizados) (ver también Sección 15.1.2).

Es probable que sea útil consultar a un economista en salud cuando se realice la clasificación de los estudios económicos en salud. Esto se debe a que es posible que, cuando se realiza una inspección más cuidadosa, los estudios económicos en salud que señalan utilizar un tipo de diseño de estudio (.ej. análisis de costo-beneficio), utilicen en realidad otro tipo de estudio (p.ej. análisis de costo-efectividad). Por lo tanto, se necesita ser particularmente cuidadoso cuando se clasifican los estudios económicos encontrados durante una revisión (Zarnke 1997).

Según el alcance de la revisión crítica de estudios económicos en salud y los tipos de estudios que se consideren para su inclusión (ver Sección 15.2.3), los estudios económicos en salud se pueden excluir en este estadio debido a la clasificación del diseño del estudio. Nuevamente, los motivos para la exclusión de los estudios económicos en salud en este estadio se deben informar en las tablas “Características de los estudios excluidos”.

15.5.2 Evaluación crítica de la calidad metodológica

El siguiente estadio de la investigación es realizar una evaluación crítica de la calidad metodológica de los restantes estudios económicos en salud, con el fin de abordar el riesgo de sesgo. La variabilidad en la calidad de la realización y redacción de los análisis económicos en salud está bien documentada (Neumann 2005). El objetivo fundamental de la evaluación crítica de los estudios económicos en salud es evaluar si los mismos describen los métodos, suposiciones, modelos y posibles sesgos de una manera transparente y basada completamente en las pruebas disponibles, cuya solidez sea fácilmente accesible a cualquier lector crítico (Rennie 2000).

La evaluación crítica de los estudios económicos en salud se puede notificar mediante el uso de listas de verificación que se han desarrollado para guiar las evaluaciones de la calidad metodológica. Cuando se utilicen listas de verificación para informar la evaluación crítica de los estudios económicos en salud en una revisión Cochrane, los detalles bibliográficos de la lista de verificación se deben citar en la sección “Obtención y análisis de los datos”. Cualquiera que sea la lista de verificación que se utilice, también es útil considerar la inclusión de tablas adicionales para resumir las listas de verificación completadas para los estudios económicos en salud incluidos en la revisión publicada.

La fiabilidad de una evaluación económica completa (ver Sección 15.5.2) se basa en parte en el uso de datos de efectividad fiables, por lo que parte de la evaluación crítica de una evaluación económica completa realizada junto con un estudio de efectividad único (p.ej. un ensayo aleatorizado) incluye considerar todas las fuentes de sesgo potencial que se puedan aplicar al estudio de efectividad utilizado (ver Capítulo 8). Para este tipo de estudio de evaluación económica completa la evaluación crítica consistirá entonces de las dos partes siguientes:

1. Evaluación del riesgo de sesgo en los resultados del estudio de efectividad único en el cual se basó el estudio de evaluación económica completa, proporcionado mediante una lista de verificación reconocida para los estudios de efectividad.
2. Evaluación de la calidad metodológica del estudio de evaluación económica completa, proporcionado mediante una lista de verificación reconocida para las evaluaciones económicas realizadas junto con diseños de estudios únicos.

Se han desarrollado varias listas de verificación para guiar la evaluación crítica de los estudios económicos en salud. Aunque ninguna de ellas se ha validado formalmente, dos se han examinado con más detalle que la mayoría:

- British Medical Journal Checklist para autores y revisores pares de artículos económicos (Drummond 1996);
- Lista CHEC para al evaluación de la calidad metodológica de las evaluaciones económicas (Evers 2005).

Estas listas de verificación se reproducen en la Figura 15.5.a y la Figura 15.5.b. En las revisiones Cochrane se recomienda el uso de la "lista de verificación de Drummond" y la "lista de verificación de Evers" para describir la evaluación de la calidad metodológica de las evaluaciones económicas completas realizadas junto con estudios de efectividad únicos, y también para detallar la evaluación crítica de evaluaciones económicas parciales mediante el subgrupo de ítems aplicables de la lista de verificación (ver también Sección 15.1.2).

Si el alcance de la revisión crítica de estudios económicos en salud incluye estudios relevantes de modelado económico (ver Sección 15.2.3), las evaluaciones de la calidad metodológica de dichos estudios deberán describirse mediante una lista de verificación diferente, ya que la "lista de verificación de Drummond" y la "lista de verificación de Evers" son pertinentes pero no suficientes para estudios de modelado. Se recomienda la "lista de verificación de Phillips" para detallar la evaluación crítica de la calidad metodológica de los estudios de modelado económico (Philips 2004). El uso de esta lista de verificación se puede complementar mediante la referencia a una jerarquía publicada de fuentes de datos que establece las fuentes de los datos reconocidas como las mejores fuentes disponibles para presentar cada parámetro en un modelo económico (Cooper 2005).

Puede ser útil describir la evaluación crítica de la calidad metodológica de todos los tipos de evaluación económica completa mediante un resumen estructurado NHS EED, si está disponible, para complementar el uso de la listas de verificación (ver también Sección 15.3.2). Esto se debe a que los resúmenes estructurados NHS EED incluyen la evaluación crítica de la calidad de un estudio según las mismas dimensiones de la calidad reflejadas en las listas de verificación recomendadas anteriormente.

Hasta el momento no hay criterios metodológicos mínimos ampliamente validados para aplicarlos al cribado de los estudios económicos, para su inclusión en las revisiones sistemáticas. Por lo tanto, las decisiones de incluir o excluir dichos estudios deberán tomarse sobre la base de una evaluación general con respecto a su calidad metodológica, así como a su relevancia en términos de las preguntas económicas, las intervenciones, las poblaciones y los resultados estudiados (ver Sección 15.4.1). Los criterios de elegibilidad con respecto a las dimensiones de la calidad metodológica de los estudios económicos en salud se deben señalar en la sección "Obtención y análisis de los datos".

También es importante destacar que hasta la fecha ha habido relativamente poca investigación empírica de la repercusión de las decisiones de incluir estudios económicos que cumplan algunos, pero no todos los estándares de calidad metodológica sobre los resultados de una revisión crítica de los estudios económicos en salud. Sin embargo, al

igual que con la selección de los criterios de elegibilidad con respecto a la calidad y el diseño de los estudios de efectividad, así como al diseño de los estudios de evaluación económica en salud (ver también Sección 15.2.3), es factible que el uso de diferentes fuentes de datos para las medidas de uso de recursos, costos y costo-efectividad tenga al menos la posibilidad de repercutir sobre los resultados (ver Sección 15.7).

Figura 15.5.a: Lista de verificación de Drummond (Drummond 1996)

Item	Sí	No	No claro	No apropiado
Diseño del estudio				
1. Se señala la pregunta de investigación.				
2. Se señala la importancia económica de la pregunta de investigación.				
3. Se señala/n y justifica/n claramente el/los punto/s de vista del análisis.				
4. Se señala la justificación para seleccionar los programas o intervenciones alternativos comparados.				
5. Se describen claramente las alternativas comparadas.				
6. Se señala la forma de evaluación económica utilizada.				
7. La selección de la forma de evaluación económica se justifica con respecto a las preguntas abordadas.				
Obtención de los datos				
8. Se señala/n la/s fuente/s de las estimaciones de efectividad utilizadas.				
9. Se proporcionan detalles del diseño y los resultados del estudio de efectividad (si se basa en un estudio único).				
10. Se proporcionan detalles de los métodos de síntesis o metanálisis de las estimaciones (si se basa en una síntesis de varios estudios de efectividad).				
11. Se señalan claramente la/s medida/s de resultado primaria/s para la evaluación económica.				
12. Se señalan los métodos para evaluar los beneficios.				
13. Se proporcionan detalles de los sujetos de los cuales se obtuvieron las evaluaciones.				
14. Los cambios en la productividad (si se incluyeron) se describen por separado.				
15. Se discute la relevancia de los cambios en la productividad para la pregunta de estudio.				
16. Las cantidades del uso de recursos se describen por separado de sus costos por unidad.				
17. Se describen los métodos para la estimación de las cantidades y los costos por unidad.				

18. Se registran los datos de la moneda y los precios.		
19. Se proporcionan detalles de los ajustes de los precios según la moneda para la inflación o la conversión de la moneda.		
20. Se proporcionan detalles de cualquier modelo utilizado.		
21. Se justifica la selección del modelo utilizado y los parámetros clave sobre los cuales se basó la misma.		
Análisis e interpretación de los resultados		
22. Se señala el horizonte de tiempo de los costos y beneficios.		
23. Se señala/n la/s tasa/s de descuento.		
24. Se justifica la selección de la/s tasa/s de descuento.		
25. Se proporciona una explicación si no se descuentan los costos y los beneficios.		
26. Se proporcionan detalles de las pruebas estadísticas y los intervalos de confianza para los datos estocásticos.		
27. Se proporciona el enfoque del análisis de sensibilidad.		
28. Se justifica la selección de las variables para el análisis de sensibilidad.		
29. Se justifican los rangos en los cuales variaron las variables.		
30. Se comparan alternativas pertinentes.		
31. Se describe el análisis de incremento.		
32. Se presentan los resultados principales en forma agregada y desagregada.		
33. Se proporciona la respuesta a la pregunta de estudio.		
34. Las conclusiones proceden de los datos descritos.		
35. Las conclusiones se acompañan de recomendaciones apropiadas.		

Figura 15.5.b: Lista de verificación de Evers (Evers 2005)

Item	Sí	No
1. ¿Se describe claramente la población de estudio?		
2. ¿Se describen claramente las alternativas opuestas?		
3. ¿La pregunta de investigación bien definida se formula de una forma que se puede responder?		
4. ¿El diseño del estudio económico es apropiado para el objetivo señalado?		
5. ¿El horizonte de tiempo seleccionado es apropiado para incluir costos y consecuencias relevantes?		
6. ¿La selección de la perspectiva real es apropiada?		
7. ¿Se identifican todos los costos importantes y relevantes para cada alternativa?		
8. ¿Todos los costos se miden de forma apropiada en unidades físicas?		

9. ¿Los costos se valoran de forma apropiada?
10. ¿Se identifican todos los resultados importantes y relevantes para cada alternativa?
11. ¿Todos los resultados se miden de forma apropiada?
12. ¿Todos los resultados se valoran de forma apropiada?
13. ¿Se realiza un análisis de aumento de los costos y de los resultados de las alternativas?
14. ¿Todos los costos y resultados futuros de descuentan de forma apropiada?
15. ¿A todas las variables importantes, cuyos valores son inciertos, se les realizaron análisis de sensibilidad apropiados?
16. ¿Las conclusiones proceden de los datos descritos?
17. ¿El estudio discute la generalizabilidad de los resultados a otros ámbitos y grupos de pacientes/clientes?
18. ¿El artículo indica que no hay conflictos de interés potenciales del/los investigador/es y los patrocinador/es del estudio?
19. ¿Se discuten de forma apropiada los aspectos éticos y de distribuciones?

15.6 Análisis y presentación de los resultados

Las directrices sobre los métodos analíticos para los componentes económicos de las revisiones Cochrane hacen énfasis en la tabulación de las características y los resultados de los estudios económicos en salud incluidos. Esto se puede complementar con un resumen narrativo que se enfoque en la evaluación crítica de los estudios incluidos y la discusión de sus hallazgos principales. Adicionalmente, en algunas circunstancias se puede considerar realizar un metanálisis de los datos de uso de recursos o los costos, o desarrollar un modelo económico. Estas opciones se describen con más detalle en las secciones siguientes. Las opciones adicionales para el análisis de los estudios económicos en salud y la presentación de los resultados de dichos análisis se deben evaluar mediante otras investigaciones metodológicas (ver Sección 15.9).

15.6.1 Presentación de los resultados en tablas

En una revisión Cochrane, las tablas “Características de los estudios incluidos” proporcionan un espacio natural para presentar detalles de las características de los estudios económicos en salud incluidos, como el año del estudio, los detalles de las intervenciones y los comparadores, el diseño del estudio, las fuentes de datos, la jurisdicción y el ámbito, la perspectiva analítica y el horizonte de tiempo (ver también Sección 15.4.2). Los revisores pueden considerar incluir tablas adicionales para resumir las listas de verificación completadas para describir las evaluaciones de la calidad metodológica de los estudios económicos en salud incluidos (ver también Sección 15.5.2).

Los resultados de los estudios económicos en salud incluidos se pueden resumir mediante las tablas “Características de los estudios incluidos”, las tablas adicionales o ambas. En cualquier caso, de ser posible se deben presentar las estimaciones puntuales de las medidas de los ítems de uso de recursos o costos, con las medidas de incertidumbre asociadas para la intervención objetivo y para cada uno de sus comparadores, así como las estimaciones puntuales de *aumento* de los costos o costo-efectividad, nuevamente con

sus medidas de incertidumbre asociadas. También es importante señalar el estado de la moneda y el precio del año, junto con las estimaciones de los costos o el aumento de los costos (si se describen).

Pudiera ser posible convertir las estimaciones de costos a una moneda y a un precio del año comunes, con el objetivo de facilitar la comparación de las estimaciones obtenidas de diferentes estudios. Se debe utilizar una tasa de intercambio internacional basada en las Paridades de Poder de Compra (en inglés, *Purchasing Power Parities* [PPP]) para convertir las estimaciones de costos a la moneda objetivo, y la medida deflacionista producto interno bruto (PBI) (o medidas deflacionistas de precios específicas para el PBI) para convertir las estimaciones de costos a un precio del año fijo. Los grupos de datos que contienen las tasas de conversión de la PPP y los valores de las medidas deflacionistas del PBI están disponibles en la International Monetary Fund in the World Economic Outlook Database (actualizada cada dos años: ver www.imf.org/external/data.htm). La conversión de las estimaciones de costos a una moneda y a un precio del año comunes sólo se deben realizar previa consulta con un economista en salud experto. El GCCME intentará producir directrices metodológicas adicionales sobre este tema en su momento.

15.6.2 Resumen narrativo de los resultados

Las revisiones Cochrane pueden incluir resúmenes narrativos de las características y los resultados principales de los estudios económicos incluidos, por ejemplo, medidas de aumento del uso de recursos, costo y costo-efectividad, para complementar y proporcionar comentarios sobre los resultados tabulados. Esto se puede localizar en la sección “Resultados”, junto con un resumen narrativo de los resultados de los estudios de efectividad (ver Capítulo 11, Sección 11.7).

El objetivo fundamental del resumen narrativo es hacer explícito para el usuario final el grado en el cual las estimaciones de costos y uso de recursos obtenidas de múltiples estudios son homogéneas entre los estudios. Esto se puede lograr mediante la descripción de las diferencias en los métodos para la evaluación y los patrones del uso de recursos y los costos entre los grupos de comparación, dentro de los estudios incluidos y entre ellos, con explicaciones potenciales para cualquier inconsistencia en los resultados entre los estudios. Como se discute anteriormente en este capítulo, los estudios de evaluación económica se construyen de forma diferente y para diferentes fines (ver también Sección 15.1.2). Este es el único factor que puede dar lugar a heterogeneidad entre los estudios en sus métodos y resultados. Cuando existe heterogeneidad entre los estudios económicos en sus métodos o resultados, llamar la atención sobre estas fuentes potenciales de heterogeneidad *estadística* puede ayudar a resumir la bibliografía económica internacional de una forma explícita, que es probable que sea útil para los usuarios finales de las revisiones (Gilbody 1999). Es importante evitar el uso de esta sección como una forma de análisis que dé lugar a recomendaciones sobre costo-efectividad (ver también Sección 15.8).

Otras características de buena práctica en un resumen narrativo de estudios económicos en salud incluidos son:

- describir el número general de estudios económicos en salud seleccionados para inclusión en la revisión, según el diseño del estudio;
- esbozar las preguntas económicas abordadas en los estudios incluidos;
- describir los diseños de los estudios incluidos;
- describir los puntos de vista analíticos adoptados en los estudios incluidos;
- describir los horizontes de tiempo adoptados en los estudios incluidos;
- discutir las medidas de aumento del uso de recursos, costos y costo-efectividad informadas en los estudios incluidos;

- describir las medidas de incertidumbre junto con las medidas de uso de recursos, costos y costo-efectividad extraídas de los estudios incluidos;
- describir la moneda y los precios del año junto con las estimaciones de los costos extraídos de los estudios incluidos;
- ajustar las estimaciones de costos extraídas de cada estudio incluido a una moneda y un precio del año comunes, de ser posible;
- destacar las características clave de los análisis de sensibilidad realizados y la consistencia de los resultados, dentro de los análisis de sensibilidad y entre los estudios incluidos;
- discutir la calidad metodológica general y las limitaciones de los estudios incluidos;
- discutir la relevancia y generalizabilidad de los resultados de los estudios incluidos a otras jurisdicciones y ámbitos; y
- discutir la calidad de los datos de efectividad utilizados en los estudios económicos en salud incluidos y la relación entre los resultados utilizados y las estimaciones en el componente de efectividad de la revisión Cochrane.

Una opción adicional es proporcionar vínculos a los resúmenes NHS EED u otros resúmenes estructurados de estudios de evaluación económica completa, de estar disponibles. Los resúmenes NHS EED estructurados incluyen información sobre las características y los resultados de las evaluaciones económicas en salud completas (ver también Sección 15.3.2). Algunas revisiones sistemáticas contienen resúmenes NHS EED de evaluaciones económicas completas incluidas en un apéndice, así como un resumen narrativo de los resúmenes en el texto principal de la revisión (Rodgers 2006, Fayter 2007).

15.6.3 Metanálisis de los datos de uso de recursos y costos

Actualmente no hay acuerdo acerca de los métodos para combinar las estimaciones combinadas de costo-efectividad (p.ej. aumento de la proporción de costo-efectividad, costo-utilidad o costo-beneficio), extraídas de múltiples evaluaciones económicas, mediante un metanálisis u otros métodos de síntesis cuantitativa. Sin embargo, en principio, si las estimaciones de las medidas de uso de recursos y los costos para una intervención y su comparador están disponibles en una magnitud común (y sus medidas de incertidumbre asociadas) para dos o más estudios incluidos, se pueden combinar mediante un metanálisis. En la práctica se recomienda extrema precaución cuando se considera realizar un metanálisis de los datos de uso de recursos o costos como parte de una revisión Cochrane. Antes de tomar cualquier decisión de combinar las estimaciones mediante un metanálisis, se debe prestar atención particular a si la medida en cuestión tiene un significado equivalente entre los estudios.

El uso de recursos y los costos son sensibles a la variabilidad entre los ámbitos *dentro* de un país y *entre* los países, en cuanto a las características del contexto local como los precios locales o aspectos de la organización y administración de los servicios (Drummond 2001, Sculpher 2004). Esto puede limitar la generalizabilidad y la transferencia de estimaciones de costos, uso de recursos y, por implicación, las estimaciones de costo-efectividad, entre los ámbitos. También es el motivo principal de que los datos de uso de recursos y los costos relacionados con poblaciones objetivo específicas y jurisdicciones de interés se consideren como la mejor fuente disponible de datos para las evaluaciones económicas, para utilizarlos en los procesos de toma de decisiones en la asignación de recursos en el ámbito específico (Cooper 2005). Estos aspectos han generado debate sobre si el metanálisis de las medidas de uso de recurso o costos entre límites geográficos o políticos más amplios puede generar resultados significativos, cómo se deben interpretar los resultados de los metanálisis y qué valor adicional pueden tener los resultados para los usuarios finales de las revisiones Cochrane. Discusiones adicionales acerca de los aspectos de la aplicabilidad y la transferencia de evaluaciones económicas en salud también se

pueden encontrar en los textos de Hutubessy y cols. y Kumaranayake y Walker (Kumaranayake 2002, Hutubessy 2003).

Por otra parte, la interrogante de si las estimaciones específicas del uso de recursos o los costos son generalizables o transferibles entre los ámbitos se puede considerar una pregunta empírica. En circunstancias en las que existen pruebas de poca variación en el uso de recursos o los costos entre los estudios, se puede considerar legítimo presentar una estimación combinada. Por otra parte es importante que la distribución de los costos se presente con claridad. Muchas revisiones Cochrane completadas incluyen metanálisis de datos de uso de recursos. Un pequeño número de revisiones Cochrane incluyen metanálisis de datos de costo, aunque éstas no siempre se acompañan de una evaluación crítica de los métodos utilizados para generar dichos datos.

Si en una revisión Cochrane se realizan metanálisis de los datos de uso de recursos o costos, siempre deben estar apoyados por una evaluación crítica cuidadosa de los métodos utilizados para derivar dichas estimaciones dentro de los estudios económicos en salud correspondientes (ver Secciones 15.5.2 y 15.6.2), junto con el uso de métodos estadísticos para investigar e incorporar la heterogeneidad entre estudios (p.ej. I^2, ji cuadrado, modelos de efectos aleatorizados; ver Capítulo 9, Sección 9.5). Las estimaciones de costos obtenidas de múltiples estudios se deben ajustar a una moneda y un precio del año comunes antes de combinar los datos (ver también Sección 15.6.1). Los revisores deben consultar el Capítulo 9 para recomendaciones adicionales sobre los procedimientos estadísticos que fundamentan el metanálisis.

Si se realizan metanálisis de datos de uso de recursos o costos, se debe incluir un resumen narrativo en la sección “Resultados” para comentar la dirección y la magnitud de los resultados, así como su precisión. De manera similar, si se incluyen dos o más estudios económicos en salud en una revisión, pero se decide no combinar (en un metanálisis) los datos de uso de recursos o costos que se obtuvieron de estos estudios, esta decisión se puede señalar en la sección “Métodos” (ver Recuadro 15.6.a para un ejemplo de este tipo de declaración).

Recuadro 15.6.a: Declaración de una decisión de no realizar un metanálisis de datos de uso de recursos o costos

“[Los resultados del uso de recursos y costos] no se combinaron porque se consideró que los resultados no eran comparables entre los ensayos… Los resultados son específicos para los países en los cuales se realizaron los estudios debido a diferencias entre los sistemas de salud pública. Las publicaciones detalladas muestran una distribución muy diferente de los costos entre diferentes ítems en diferentes países.” (Birks 2006).

15.6.4 Desarrollo de un modelo económico

Las revisiones Cochrane pueden contribuir con componentes clave de la evidencia requerida para desarrollar una evaluación económica completa posterior o paralela, incluido el uso de un enfoque de análisis de decisiones para la combinación o el modelado de las pruebas disponibles sobre los costos y los efectos de la intervención (ver también Secciones 15.1.2 y 15.1.3). Este enfoque incluye habitualmente la determinación de la estimación puntual, y la descripción de la distribución conjunta, del aumento de los costos y efectos debido a una intervención (en términos de costo-efectividad, costo-utilidad o costo-beneficio), en comparación con una alternativa relevante, en una población y un ámbito definidos, y con los costos y los resultados incluidos que se acordó eran relevantes

desde un punto de vista analítico específico (p.ej. paciente, proveedor de atención sanitaria o un tercero [financiador, sistema de salud, sociedad]).

Los métodos de modelado económico no se analizan en detalle aquí, ya que no se recomienda su uso habitual como parte del proceso de una revisión Cochrane. Sin embargo, a los autores de revisiones Cochrane que deseen abordar los aspectos económicos de las intervenciones "en profundidad", se les recomienda que colaboren con investigadores con experiencia en el desarrollo de modelos económicos. En ocasiones es posible desarrollar una estructura general para un modelo económico como parte de una revisión Cochrane, en la que las entradas y las salidas básicas del modelo son similares entre diferentes ámbitos, pero en los que algunos (o incluso todos) los datos que se necesitan para completar el modelo son específicos para un ámbito local.

Además, a pesar de los aspectos ya discutidos con respecto a la generalizabilidad y la transferencia de los resultados de las evaluaciones económicas entre jurisdicciones y ámbitos (ver Sección 15.6.3), no es posible descartar que algunas veces se puede considerar válido (aunque necesita tiempo, recursos y experiencia) desarrollar uno o más modelos económicos para su publicación en una revisión Cochrane. Por ejemplo, una motivación para desarrollar un modelo económico como parte de una revisión Cochrane, puede ser la intención de utilizar la revisión para informar directamente el diseño de investigaciones futuras que incorporarán un componente de evaluación económica. En estas circunstancias el desarrollo de un modelo puede ayudar a aclarar las suposiciones estructurales y los parámetros que se deben considerar en una evaluación económica, y los datos que será necesario obtener durante la investigación. Si se sigue este tipo de enfoque en una revisión Cochrane, se debe aclarar que cada modelo económico ejemplo tiene como objetivo proporcionar una evaluación ilustrativa de costo-efectividad de las intervenciones que se comparan, en una jurisdicción ejemplo y en un punto temporal determinado.

También se les recomienda a quienes desarrollan modelos económicos que consideren utilizar la evidencia presente en las revisiones Cochrane para informar el desarrollo de modelos económicos. Los esfuerzos para incorporar la evidencia económica en las revisiones Cochrane mediante los métodos esbozados en este capítulo se dirigen en parte a aumentar la relevancia y la aplicabilidad de las revisiones Cochrane para su uso en ejercicios de modelado de evaluaciones económicas completas posteriores o paralelas.

15.7 Análisis de los sesgos de notificación

Se ha reconocido ampliamente que las presiones comerciales y de otro tipo pueden afectar el financiamiento de los estudios y la publicación de los resultados de los estudios centrados en el valor económico de las intervenciones en la atención sanitaria (Drummond 1992). A pesar de esto, hasta hace poco tiempo se han realizado relativamente pocas investigaciones centradas en el aspecto del sesgo de publicación y los sesgos relacionados en los estudios de evaluación económica, en comparación con la cobertura de este aspecto en los estudios de efectividad. Sin embargo, varios estudios recientes han comenzado a examinar este tema mediante revisiones sistemáticas y métodos de síntesis de la investigación.

Bell y colegas realizaron una revisión sistemática de los estudios de costo-efectividad en la atención sanitaria publicados, y encontraron que fue más probable que los estudios patrocinados por la industria describieran tasas por debajo o muy cercanas a los umbrales de aceptabilidad de costo-efectividad propuestos con mayor frecuencia, cuando se comparan con los estudios patrocinados por otras fuentes (Bell 2006). Miners y colegas realizaron una revisión sistemática para comparar la evidencia en costo-efectividad

enviada al *National Institute of Health and Clinical Excellence* (NICE) por los fabricantes de tecnologías relevantes de atención sanitaria y por grupos universitarios de evaluación contratados, respectivamente (Miners 2005). Este estudio encontró que la estimación del aumento de las tasas de costo-efectividad enviadas por los fabricantes fueron, como promedio, significativamente menores que las proporcionadas por los grupos de evaluación para la misma tecnología. Friedberg y colegas encontraron que fue 1/8 más probable que los análisis económicos publicados de nuevos fármacos utilizados en oncología financiados por compañías farmacéuticas, alcanzaran conclusiones cuantitativas desfavorables (y 1,4 veces más probable que alcanzaran conclusiones cualitativas favorables) cuando se compararon con estudios financiados sin fines de lucro (Friedberg 1999). Otras revisiones que se centran en este aspecto han establecido conclusiones muy similares (Freemantle 1997, Azimi 1998, Lexchin 2003). Un tema de discusión frecuente en estos estudios de revisión metodológica es la sospecha de los autores de que es probable que los sesgos de notificación o publicación en los patrones de resultados observados sean instrumentales. La hipótesis general es que los análisis económicos con resultados que indican que una intervención puede ser económicamente poco atractiva, no son publicados conciente o inconcientemente, por los patrocinadores, los autores ni los editores de las revistas.

Sin embargo, todos los estudios de revisión metodológica anteriores están limitados por su diseño (habitualmente los autores reconocen y discuten dichas limitaciones). El diseño de estudio ideal y más sólido para investigar la presencia de sesgos de notificación y publicación incluiría la comparación directa de los hallazgos publicados y no publicados dentro de los estudios, o la comparación directa de los hallazgos de estudios publicados y no publicados (Song 2000). Como tal, es claramente difícil lograr una comparación sistemática y exhaustiva debido a las dificultades inherentes para identificar todos los análisis económicos relevantes no publicados. A falta de dichos datos, no es posible descartar explicaciones alternativas para los patrones de resultados observados (p.ej. los resultados pudieran reflejar las verdaderas distribuciones de las tasas de aumento de costo-efectividad).

En el Capítulo 10 se presentan los métodos para abordar el sesgo de publicación en las revisiones sistemáticas que se pueden aplicar con las mismas advertencias en las revisiones sistemáticas de estudios económicos. Las propuestas que se han indicado para ayudar a abordar el sesgo de publicación y los sesgos relacionados en los estudios de evaluación económica, como los que se pueden encontrar en las revisiones Cochrane, son:

1. estimular un enfoque consistente y más transparente para la realización y publicación de los análisis económicos, mediante la promulgación de guías de buena práctica y listas de verificación para su uso en la evaluación critica de dichos estudios, en particular los estudios basados en revisiones y los estudios de modelado;
2. aumentar el examen minucioso de los envíos a las revistas en busca de posibles conflictos de interés de los patrocinadores y los autores del estudio; y
3. incrementar el acceso a todos los datos subyacentes utilizados en una evaluación económica con el fin de aumentar la transparencia de los métodos.

15.8 Interpretación de los resultados

La interpretación de los resultados de una revisión sobre estudios económicos en salud depende de las preguntas económicas específicas y el contexto de relevancia para una decisión determinada con respecto a la provisión de atención sanitaria. En las revisiones Cochrane (dirigidas a una audiencia internacional) hay un gran número de posibles preguntas económicas y factores contextuales que las diferentes instituciones que toman decisiones deben tener en cuenta. Debido a este contexto global, sencillamente no es

factible interpretar los resultados de una revisión crítica de múltiples estudios de evaluación económica con el objetivo de establecer conclusiones acerca de la adopción o el rechazo de, por ejemplo, un tratamiento o una prueba diagnóstica en salud. Sin embargo, aunque en estas circunstancias es poco probable que la revisión Cochrane proporcione el aspecto central de cualquier evaluación de política, aún puede ayudar a refinar una discusión económica y colocarla en un contexto internacional (Gilbody 1999).

En un tema de revisión con pocos, o no relevantes, estudios de evaluación económica de alta calidad, la revisión crítica de los estudios económicos en salud puede servir para destacar una falta de pruebas económicas que las investigaciones futuras pudieran tener que abordar. La necesidad de estudios de evaluación económica adicionales se debe señalar dentro del apartado "Implicaciones para la investigación" en la sección "Conclusiones de los autores" de la revisión. El Recuadro 15.8.a muestra dos ejemplos de este tipo de declaración. También se debe considerar que como una evaluación económica completa se basa en la disponibilidad de datos fiables sobre la efectividad de la intervención, la falta de estudios de efectividad sólidos repercutiría claramente en la factibilidad y disponibilidad de estudios de evaluación económica completa. Nuevamente, aunque las revisiones Cochrane y de otro tipo no pueden sortear esta limitación, deben llamar la atención sobre ella en la sección "Conclusiones".

Recuadro 15.8.a: Se destaca la necesidad de estudios económicos adicionales en las conclusiones

"La mayoría de las veces el costo de la intervención no se calcula [en los estudios incluidos]. Esta información es fundamental. En los estudios futuros se debe calcular el ahorro de los costos y equilibrarlo contra los costos potenciales de la intervención… La pregunta de si es posible proporcionar servicios costo-efectivos es crítica para el medio de atención de salud actual. Por lo tanto, se necesitan estudios que midan el costo, así como los efectos de las intervenciones farmacológicas." (Beney 2000).

15.9 Conclusiones

Este capítulo ha esbozado un marco metodológico para incorporar la evidencia de los estudios económicos en salud al proceso de las revisiones Cochrane. Aunque es muy poco probable ni recomendable que este ejercicio produzca declaraciones acerca de si "la intervención X es costo-efectiva", puede ayudar a quienes toman las decisiones a comprender la estructura del problema de asignación de recursos que ellos enfrentan, los principales parámetros que es necesario considerar, la variación entre los ámbitos en cuanto a uso de recursos, costos y costo-efectividad, y los posibles motivos para estas variaciones (Drummond 2002). La incorporación de la evidencia económica también puede mejorar la utilidad y aplicabilidad de las revisiones Cochrane como una fuente de datos para evaluaciones económicas completas posteriores (o paralelas). Es de prever que estas directrices se refinen y actualicen debido a que están sujetas a la crítica adicional de una audiencia más amplia, y que los métodos se desarrollen aún más gracias a la experiencia con su uso en revisiones Cochrane y en investigaciones metodológicas futuras.

El proceso de desarrollo de estas directrices también ha ayudado a aclarar prioridades clave para las investigaciones futuras dirigidas a desarrollar y probar métodos alternativos para la identificación, la evaluación, el análisis y la presentación de la evidencia sobre los aspectos económicos de las intervenciones. Las prioridades clave de investigación incluyen: el desarrollo adicional de un enfoque equilibrado para resumir los resultados de los componentes económicos de las revisiones, la evaluación de la repercusión sobre los resultados de las revisiones económicas y de la aplicación de diferentes criterios o

umbrales de calidad metodológica para la inclusión de estudios de evaluación económica. Otros aspectos a considerar son la evaluación de los métodos que utilizan datos a nivel del paciente para investigar y tratar la heterogeneidad entre los ámbitos en el uso de recursos, los costos y las utilidades (y otras medidas de preferencia para los estados de salud). Éstas y otras prioridades en la investigación de los métodos se enumeran en las páginas "Investigación" del sitio web del GCCME (ver Recuadro 15.10.a).

15.10 Información del capítulo

Autores: Ian Shemilt, Miranda Mugford, Sarah Byford, Michael Drummond, Eric Eisenstein, Martin Knapp, Jacqueline Mallender, David McDaid, Luke Vale, Damian Walker del Grupo Campbell y Cochrane de Métodos Económicos.

La versión en inglés de este capítulo se debería citar como: Shemilt I, Mugford M, Byford S, Drummond M, Eisenstein E, Knapp M, Mallender J, McDaid D, Vale L, Walker D. Chapter 15: Incorporating economics evidence. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: Dawn Craig, Julian Higgins, Kevin Marsh y John Nixon por sus comentarios sobre las versiones anteriores.

Recuadro 15.10.a: El Grupo Campbell y Cochrane de Métodos Económicos

El Grupo Campbell y Cochrane de Métodos Económicos (GCCME) se registró formalmente como un grupo de métodos de la Colaboración Cochrane en 1998 y se registró conjuntamente como un grupo de métodos de la Colaboración Campbell desde 2004. Los objetivos fundamentales del grupo incluyen dentro de los recursos disponibles, los siguientes:

- promover y apoyar la consideración de los aspectos económicos dentro de las revisiones sistemáticas;
- desarrollar métodos económicos para las revisiones Cochrane que sean relevantes para los consumidores de las revisiones y que sean apropiados, no sesgados y objetivos en términos de su aplicación; y
- vincular a los revisores y editores con economistas que puedan ayudar con las revisiones o proporcionen asesoría especializada y revisión externa por expertos.

Muchas revisiones Cochrane ya incluyen aspectos económicos de las intervenciones. Sin embargo, con este capítulo es la primera vez que se incluyen en el *Manual* recomendaciones detalladas sobre el uso de métodos económicos en las revisiones Cochrane. Las versiones futuras de este capítulo se informarán mediante un programa en curso de investigaciones metodológicas y experiencia adicional de las revisiones Cochrane que incorporen pruebas económicas.
E-mail: research@c-cemg.org
Sitio web: www.c-cemg.org

15.11 Referencias

Aguilar-Ibañez 2005
Aguiar-Ibañez R, Nixon J, Glanville J, Craig D, Rice S, Christie J, Drummond MF. Economic evaluation databases as an aid to healthcare decision-makers and researchers. *Expert Review of Pharmacoeconomics and Outcomes Research* 2005; 5: 721-722.

Azimi 1998
Azimi NA, Welch HG. The effectiveness of cost-effectiveness analysis in containing costs. *Journal of General Internal Medicine* 1998; 13: 664-669.

Bell 2006
Bell CM, Urbach DR, Ray JG, Bayoumi A, Rosen AB, Greenberg D, Neumann PJ. Bias in published cost effectiveness studies: systematic review. *BMJ* 2006; 332: 699-703.

Beney 2000
Beney J, Bero LA, Bond C. Expanding the roles of outpatient pharmacists: effects on health services utilisation, costs, and patient outcomes. *Cochrane Database of Systematic Reviews* 2000, Issue 3. Art No: CD000336.

Best 2000
Best L, Simmonds P, Baughan C, Buchanan R, Davis C, Fentiman I, George S, Gosney M, Northover J, Williams C, Colorectal Meta-analysis Collaboration. Palliative chemotherapy for advanced or metastatic colorectal cancer. *Cochrane Database of Systematic Reviews* 2000, Issue 2. Art No: CD001545.

Birks 2006
Birks J, Harvey RJ. Donepezil for dementia due to Alzheimer's disease. *Cochrane Database of Systematic Reviews* 2006, Issue 1. Art No: CD001190.

Briggs 2006
Briggs A, Sculpher M, Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford (UK): Oxford University Press, 2006.

Brown 2007
Brown SR, Nelson RL. Surgery for faecal incontinence in adults. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art No: CD001757.

Campbell 2007
Campbell H, Briggs A, Buxton M, Kim L, Thompson S. The credibility of health economic models for health policy decision-making: the case of population screening for abdominal aortic aneurysm. *Journal of Health Services Research and Policy* 2007; 12: 11-17.

Cochrane 1972
Cochrane AL. *Effectiveness and Efficiency: Random Reflections on Health Services*. London (UK): Nuffield Provincial Hospitals Trust, 1972.

Cooper 2005
Cooper N, Coyle D, Abrams K, Mugford M, Sutton A. Use of evidence in decision models: an appraisal of health technology assessments in the UK since 1997. *Journal of Health Services Research and Policy* 2005; 10: 245-250.

Craig 2007
Craig D, Rice S. *CRD Report 6: NHS Economic Evaluation Database Handbook* (3rd edition). York (UK): Centre for Reviews and Dissemination, University of York, 2007.

Donaldson 2002
Donaldson C, Mugford M, Vale L. From effectiveness to efficiency: an introduction to evidence-based health economics. In: Donaldson C, Mugford M, Vale L (editors). *Evidence-based Health Economics: From Effectiveness to Efficiency in Systematic Reviews*. London (UK): BMJ Books, 2002.

Drummond 1992
Drummond MF. Economic evaluation of pharmaceuticals: science or marketing? *Pharmacoeconomics* 1992; 1: 8-13.

Drummond 1996
Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996; 313: 275-283.

Drummond 2001
Drummond M, Pang F. Transferability of economic evaluation results. In: Drummond M, McGuire A (editors). *Economic Evaluation in Health Care: Merging Theory with Practice*. New York (NY): Oxford University Press, 2001.

Drummond 2002
Drummond M. Evidence-based medicine meets economic evaluation – an agenda for research. In: Donaldson C, Mugford M, Vale L (editors). *Evidence-based Health Economics: From Effectiveness to Efficiency in Systematic Reviews*. London (UK): BMJ Books, 2002.

Drummond 2005
Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes* (3rd edition). Oxford (UK): Oxford University Press, 2005.

Evers 2005
Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *International Journal of Technology Assessment in Health Care* 2005; 21: 240-245.

Fayter 2007
Fayter D, Nixon J, Hartley S, Rithalia A, Butler G, Rudolf M, Glasziou P, Bland M, Stirk L, Westwood M. A systematic review of the routine monitoring of growth in children of primary school age to identify growth-related conditions. *Health Technology Assessment* 2007; 11: 22.

Freemantle 1997
Freemantle N, Mason J. Publication bias in clinical trials and economic analyses. *Pharmacoeconomics* 1997; 12: 10-16.

Friedberg 1999
Friedberg M, Saffran B, Stinson TJ, Nelson W, Bennett CL. Evaluation of conflict of interest in economic analyses of new drugs used in oncology. *JAMA* 1999; 282: 1453-1457.

Gamble 2006
Gamble CL, Williamson PR, Marson AG. Lamotrigine versus carbamazepine monotherapy for epilepsy. *Cochrane Database of Systematic Reviews* 2006, Issue 1. Art No: CD001031.

Gilbody 1999
Gilbody SM, Petticrew M. Rational decision-making in mental health: the role of systematic reviews. *Journal of Mental Health Policy and Economics* 1999; 2: 99-106.

Hutubessy 2003
Hutubessy R, Chisholm D, Edejer TT. Generalized cost-effectiveness analysis for national-level priority-setting in the health sector. *Cost Effectiveness and Resource Allocation* 2003; 1: 8.

Kumaranayake 2002
Kumaranayake L, Walker D. Cost-effectiveness analysis and priority setting: Global approach without local meaning? In: Lee K, Buse K, Fustukian S (editors). *Health Policy in a Globalising World*. Cambridge (UK): Cambridge University Press, 2002.

Lavis 2005
Lavis J, Davies H, Oxman A, Denis JL, Golden-Biddle K, Ferlie E. Towards systematic reviews that inform health care management and policy-making. *Journal of Health Services Research and Policy* 2005; 10 Suppl 1: 35-48.

Lexchin 2003
Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003; 326: 1167-1170.

Maynard 2000
Maynard A, Kanavos P. Health economics: an evolving paradigm. *Health Economics* 2000; 9: 183-190.

Miners 2005
Miners AH, Garau M, Fidan D, Fischer AJ. Comparing estimates of cost effectiveness submitted to the National Institute for Clinical Excellence (NICE) by different organisations: retrospective study. *BMJ* 2005; 330: 65.

Napper 2005
Napper M, Varney J. Etext on Health Technology Assessment (HTA) Information Resources. Chapter 11: Health Economics Information. Available from: http://www.nlm.nih.gov/archive//2060905/nichsr/ehta/chapter11.html (accessed 1 January 2008).

Neumann 2005
Neumann PJ, Greenberg D, Olchanski NV, Stone PW, Rosen AB. Growth and quality of the cost-utility literature, 1976-2001. *Value in Health* 2005; 8: 3-9.

Philips 2004
Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, Woolacoot N, Glanville J. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technology Assessment* 2004; 8: 36.

Rennie 2000
Rennie D, Luft HS. Pharmacoeconomic analyses: making them transparent, making them credible. *JAMA* 2000; 283: 2158-2160.

Rodgers 2006
Rodgers M, Nixon J, Hempel S, Aho T, Kelly J, Neal D, Duffy S, Ritchie G, Kleijnen J, Westwood M. Diagnostic tests and algorithms used in the investigation of haematuria: systematic reviews and economic evaluation. *Health Technology Assessment* 2006; 10: 18.

Samuelson 2005
Samuelson PA, Nordhaus WD. *Economics*. London (UK): McGraw-Hill, 2005.

Sculpher 2004
Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, Davies LM, Eastwood A. Generalisability in economic evaluation studies in healthcare: a review and case studies. *Health Technology Assessment* 2004; 8: 49.

Shemilt 2007
Shemilt I, Mugford M, Byford S, Drummond M, Eisenstein E, Knapp M, Mallender J, McDaid D, Vale L, Walker D. Where does economics fit in? A review of economics in Cochrane Reviews. *15th Cochrane Colloquium*, Sau Paulo (Brazil), 2007.

Song 2000
Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technology Assessment* 2000; 4: 10.

Weinstein 2003
Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, Luce BR. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices - Modeling studies. *Value in Health* 2003; 6: 9-17.

Williams 1987
Williams A. Health economics: The cheerful face of the dismal science? In: Williams A (editors). *Health and Economics*. London (UK): Macmillan, 1987.

Zarnke 1997
Zarnke KB, Levine MA, O'Brien BJ. Cost-benefit analyses in the health-care literature: don't judge a study by its label. *Journal of Clinical Epidemiology* 1997; 50: 813-822.

ÍNDICE

CAPÍTULO 16: TEMAS ESPECIALES EN ESTADÍSTICA

PUNTOS CLAVE

16.1 DATOS FALTANTES

16.1.1 Tipos de datos faltantes
Tabla 16.1.a: Tipos de datos faltantes en un metanálisis
16.1.2 Principios generales para tratar los datos faltantes
16.1.3 Desviaciones estándar faltantes
16.1.3.1 Imputación de las desviaciones estándar
16.1.3.2 Imputación de las desviaciones estándar para los cambios a partir del valor inicial

16.2 ASPECTOS DEL ANÁLISIS POR INTENCIÓN DE TRATAR

16.2.1 Introducción
16.2.2 Aspectos del análisis por intención de tratar para los datos dicotómicos
16.2.3 Aspectos del análisis por intención de tratar para los datos continuos
16.2.4 Resultados condicionales disponibles solamente para subgrupos de participantes

16.3 ENSAYOS ALEATORIZADOS POR GRUPO

16.3.1 Introducción
16.3.2 Evaluación del riesgo de sesgo en los ensayos aleatorizados por grupo
16.3.3 Métodos de análisis para los ensayos aleatorizados por grupo
16.3.4 Análisis aproximados de los ensayos aleatorizados por grupo para un metanálisis: tamaños de muestra efectivos
16.3.5 Ejemplo de incorporación de un ensayo aleatorizado por grupo
16.3.6 Análisis aproximados de los ensayos aleatorizados por grupo para un metanálisis: inflación de los errores estándar
16.3.7 Aspectos de la incorporación de ensayos aleatorizados por grupo
16.3.8 Ensayos aleatorizados individuales con agrupamiento

16.4 ENSAYOS CRUZADOS

16.4.1 Introducción
16.4.2 Evaluación de la conveniencia de los ensayos cruzados
16.4.3 Evaluación del sesgo en los ensayos cruzados
16.4.4 Métodos de análisis para los ensayos cruzados
16.4.5 Métodos para incorporar los ensayos cruzados en un metanálisis
16.4.6 Análisis aproximados de los ensayos cruzados para un metanálisis
Introducción
Tabla 16.4.a: Algunos datos posibles disponibles en la publicación de un ensayo cruzado
16.4.6.1 Diferencias de medias
16.4.6.2 Diferencia de medias estandarizada
16.4.6.3 Imputación de los coeficientes de correlación
16.4.6.4 Ejemplo
16.4.7 Problemas en la incorporación de los ensayos cruzados

16.5 ESTUDIOS CON MÁS DE DOS GRUPOS DE INTERVENCIÓN

16.5.1 Introducción
16.5.2 Determinación de qué grupos de intervención son relevantes
16.5.3 Evaluación del riesgo de sesgo en los estudios con más de dos grupos
16.5.4 Cómo incluir múltiples grupos de un estudio
16.5.5 Consideraciones sobre la heterogeneidad con estudios de intervenciones múltiples
16.5.6 Ensayos factoriales

16.6 COMPARACIONES INDIRECTAS Y METANÁLISIS DE TRATAMIENTOS MÚLTIPLES

16.6.1 Introducción
16.6.2 Comparaciones indirectas
16.6.3 Metanálisis de tratamientos múltiples

16.7 MULTIPLICIDAD Y FUNCIÓN DEL AZAR

16.7.1 Introducción
16.7.2 Multiplicidad en las revisiones sistemáticas

16.8 ENFOQUES BAYESIANOS Y JERÁRQUICOS PARA EL METANÁLISIS

16.8.1 Métodos bayesianos
16.8.2 Modelos jerárquicos

16.9 EVENTOS POCO FRECUENTES (INCLUIDAS LAS FRECUENCIAS CERO)

16.9.1 Metanálisis de eventos poco frecuentes
16.9.2 Estudios con recuentos de celdas cero
16.9.3 Estudios sin eventos
16.9.4 Intervalos de confianza cuando no se observan eventos
16.9.5 Validez de los métodos de metanálisis para los eventos poco frecuentes

16.10 INFORMACIÓN DEL CAPÍTULO

16.11 REFERENCIAS

CAPÍTULO 16: TEMAS ESPECIALES EN ESTADÍSTICA

Editores: Julian PT Higgins, Jonathan J Deeks y Douglas G Altman del Grupo Cochrane de Métodos Estadísticos

Puntos clave

- Cuando los datos faltantes impiden que un estudio se incluya en un metanálisis (y los intentos de obtener los datos de los investigadores originales han sido infructuosos), cualquier estrategia para imputar los datos se debería describir y evaluar en análisis de sensibilidad.
- Los diseños no estándar, como los ensayos aleatorizados por grupos y los ensayos cruzados se deberían analizar mediante métodos apropiados al diseño. Incluso si los autores del estudio no pueden encontrar correlaciones entre los datos de resultado, a menudo los revisores pueden aplicar métodos aproximados.
- Para incluir un estudio con más de dos grupos de intervención en un metanálisis, el enfoque recomendado habitualmente es combinar los grupos pertinentes para crear una comparación pareada única.
- Las comparaciones indirectas de intervenciones pueden ser erróneas, pero hay métodos disponibles que explotan la asignación al azar, incluidas las extensiones hacia los "metanálisis de múltiples tratamientos".
- Para reducir las conclusiones erróneas debidas a los análisis estadísticos múltiples los revisores deberían señalar en el protocolo qué análisis realizarán, mantener el número de éstos en el mínimo e interpretar los hallazgos estadísticamente significativos en el contexto de cuántos análisis se realizaron.
- Los enfoques bayesianos y los modelos jerárquicos (o multinivel) permiten realizar metanálisis más complejos y pueden ofrecer algunas ventajas técnicas e interpretativas sobre los métodos estándar implementados en RevMan.
- Los estudios sin eventos no contribuyen con información acerca del riesgo relativo o del *odds ratio*. Para los eventos poco frecuentes se ha observado que el método de Peto es menos sesgado y tiene más poder estadístico que otros métodos.

16.1 Datos faltantes

16.1.1 Tipos de datos faltantes

Existen muchas fuentes potenciales de datos faltantes en una revisión sistemática o un metanálisis (ver Tabla 16.1.a). Por ejemplo, puede faltar un estudio completo en la revisión, puede faltar un resultado de un estudio, pueden faltar los datos resumen para un resultado, y puede haber participantes individuales faltantes de los datos resumen. Aquí se discuten varias fuentes potenciales de datos faltantes y se destaca en qué apartados del *Manual* hay discusiones más detalladas.

Los estudios completos pueden faltar en una revisión debido a que nunca se publicaron, se publicaron en lugares de difícil acceso, se citan con poca frecuencia o se indexaron de forma incorrecta en las bases de datos. Por lo tanto, los revisores siempre deberían estar al tanto de la posibilidad de que no hayan podido identificar estudios relevantes. Existe una fuerte posibilidad de que dichos estudios falten debido a que los hallazgos “no son interesantes” o “no son esperados” (es decir, porque existe sesgo de publicación). Este problema se discute ampliamente en el Capítulo 10. El Capítulo 6 proporciona detalles de los métodos de búsqueda exhaustiva.

Es posible que algunos estudios no proporcionen alguna información sobre los resultados de interés de la revisión. Por ejemplo, puede que no haya información sobre la calidad de vida, o sobre eventos adversos graves. A menudo es difícil determinar si esto se debería a que el resultado no se midió o a que el resultado no se proporcionó. Además, la imposibilidad de describir que los resultados se midieron puede depender de que los resultados no se proporcionaron (sesgo de notificación selectivo de resultado; ver Capítulo 8, Sección 8.13). De manera similar, pueden faltar los datos resumen para un resultado en una forma que se puedan incluir en un metanálisis. Un ejemplo frecuente son las desviaciones estándar faltantes para los resultados continuos. Con frecuencia esto es un problema cuando se investigan resultados de cambio a partir del valor inicial. La imputación de las desviaciones estándar faltantes se discute en la Sección 16.1.3. Otros ejemplos de datos resumen faltantes son los tamaños de muestra faltantes (particularmente los de cada grupo de intervención por separado), el número de eventos, los errores estándar, los tiempos de seguimiento para calcular las tasas y detalles suficientes de los resultados de tiempo hasta el evento. Los análisis inapropiados de los estudios, por ejemplo los ensayos aleatorizados por grupo o los ensayos cruzados, pueden dar lugar a datos resumen erróneos. Algunas veces es posible aproximarse a los análisis correctos de estos estudios, por ejemplo al imputar los coeficientes de correlación o las desviaciones estándar, como se discute en la Sección 16.3 para los estudios aleatorizados por grupo y en la Sección 16.4 para los ensayos cruzados. Como regla general, la mayoría de los metodólogos considera que los datos resumen faltantes (p.ej. “datos no utilizables”) no se deberían utilizar como motivo para excluir un estudio de una revisión sistemática. Es más apropiado incluir el estudio en la revisión y discutir la posible repercusión de su ausencia del metanálisis.

Es probable que en algunos, sino todos, los estudios incluidos haya individuos que falten de los resultados informados. Los análisis de los ensayos aleatorizados que no incluyen todos los participantes asignados al azar no son análisis por intención de tratar. Algunas veces es posible realizar análisis por intención de tratar, incluso si los investigadores originales no lo hicieron. En la Sección 16.2 se proporciona una discusión detallada de los aspectos del análisis por intención de tratar.

Los datos faltantes también pueden afectar los análisis de subgrupos. Si se planifican análisis de subgrupos o metarregresiones (ver Capítulo 9, Sección 9.6), se necesitan detalles de las características a nivel de estudio que diferencian a los estudios entre sí. Si las mismas no están disponibles para todos los estudios, los revisores deberían considerar solicitar más información a los autores de los estudios.

Tabla 16.1.a: Tipos de datos faltantes en un metanálisis

Tipo de datos faltantes	Algunos motivos posibles para los datos faltantes
Estudios faltantes.	Sesgo de publicación; Búsqueda no suficientemente exhaustiva.
Resultados faltantes.	Resultado no medido; Sesgo de notificación selectivo.
Datos resumen faltantes.	Sesgo de notificación selectivo; Notificación incompleta.
Individuos faltantes.	Falta de análisis por intención de tratar; Pérdidas del estudio; Sesgo de notificación selectivo.
Características faltantes a nivel de estudio (para el análisis de subgrupos o la metarregresión).	Características no medidas; Notificación incompleta.

16.1.2 Principios generales para tratar los datos faltantes

Existe una amplia bibliografía de métodos estadísticos para tratar los datos faltantes. Aquí se revisan brevemente algunos conceptos clave y se hacen algunas recomendaciones generales para los revisores Cochrane. Es importante pensar *por qué* puede haber datos faltantes. A menudo los estadísticos utilizan los términos “faltantes al azar” y “no faltantes al azar” para representar diferentes escenarios.

Se dice que los datos son “faltantes al azar” si el hecho de que sean faltantes no se relaciona con los valores reales de los datos faltantes. Por ejemplo, si algunos cuestionarios sobre calidad de vida se perdieron en el sistema postal, sería poco probable que se relacione con la calidad de vida de los participantes en el ensayo que completaron los formularios. En algunas circunstancias los estadísticos diferencian los datos “faltantes al azar” de los datos “faltantes completamente al azar”, aunque en el contexto de una revisión sistemática es poco probable que la diferencia sea importante. Es posible que los datos faltantes al azar no sean importantes. Los análisis basados en los datos disponibles tenderán a no estar sesgados, aunque se basen en un tamaño de muestra menor que el conjunto de datos originales.

Se dice que los datos son “no faltantes al azar” si el hecho de que sean faltantes se relaciona con los datos faltantes reales. Por ejemplo, en un ensayo de depresión pudiera ser poco probable que los participantes que presentan una recaída asistan a la entrevista final de seguimiento, y es más probable que tengan datos de resultado faltantes. Estos datos son “no ignorables” en el sentido de que un análisis de los datos disponibles solos habitualmente

estará sesgado. El sesgo de publicación y el sesgo de notificación selectivo dan lugar por definición a datos que son “no faltantes al azar”, lo que también ocurre a menudo con los abandonos y las exclusiones de individuos dentro de los estudios.

Las opciones principales para tratar con los datos faltantes son:

1. analizar solamente los datos disponibles (es decir, ignorar los datos faltantes);
2. imputar los datos faltantes con reemplazo de los valores, y tratarlos como si fueran datos observados (p.ej. la última observación realizada; la imputación de un resultado supuesto como si se supusiera que todos fueron resultados deficientes, la imputación de la media, la imputación basada en valores previstos de un análisis de regresión);
3. imputar los datos faltantes y tener en cuenta el hecho de se imputaron con incertidumbre (p.ej. métodos de imputación múltiple, imputación simple [como en el punto 2] con ajuste al error estándar); y
4. utilizar modelos estadísticos para permitir los datos faltantes, con supuestos acerca de su relación con los datos disponibles.

La opción 1 puede ser apropiada cuando es posible suponer que los datos son faltantes al azar. Las opciones 2 y 4 son intentos de abordar los datos no faltantes al azar. La opción 2 es práctica en la mayoría de las circunstancias y se utiliza con mucha frecuencia en las revisiones sistemáticas. Sin embargo, no reconoce la incertidumbre en los valores imputados y los resultados, generalmente, en los intervalos de confianza que son demasiado estrechos. Las opciones 3 y 4 necesitarían la participación de un estadístico experto.

A continuación se brindan cuatro recomendaciones generales para tratar con datos faltantes en las revisiones Cochrane.

- De ser posible, establecer contacto con los investigadores originales para solicitarles datos faltantes.
- Hacer explícitas las suposiciones de cualquier método utilizado para tratar los datos faltantes; por ejemplo, la suposición de que los datos eran faltantes al azar, o la suposición de que los valores faltantes tenían un valor particular como un resultado deficiente.
- Realizar análisis de sensibilidad para evaluar cuán sensibles son los resultados a los cambios razonables en las suposiciones que se hacen (ver Capítulo 9, Sección 9.7).
- Abordar en la sección “Discusión” la repercusión potencial de los datos faltantes sobre los hallazgos de la revisión.

16.1.3 Desviaciones estándar faltantes

16.1.3.1 Imputación de las desviaciones estándar

Las desviaciones estándar faltantes son una característica frecuente de los metanálisis de datos de resultados continuos. Un enfoque a este problema es imputar las desviaciones estándar. Sin embargo, antes de imputar las desviaciones estándar faltantes los revisores deberían buscar cuidadosamente estadísticos que permitan el cálculo o la estimación de la desviación estándar (p.ej. los intervalos de confianza, los errores estándar, los valores de t, los valores de p, los valores de f), como se discute en el Capítulo 7 (Sección 7.7.3).

La imputación más sencilla proviene de un valor determinado tomado de uno o más estudios. Furukawa y cols. encontraron que la imputación de las desviaciones estándar de otros estudios en el mismo metanálisis o de estudios de otros metanálisis produjo resultados aproximadamente correctos en dos estudios de casos (Furukawa 2006). Si hay varias posibles desviaciones estándar disponibles los revisores deberían decidir si utilizan el promedio, la más alta, un valor "razonablemente alto" o alguna otra estrategia. Para los metanálisis de diferencias de medias, la selección de una desviación estándar más alta disminuye la ponderación del estudio y produce un intervalo de confianza más amplio. Sin embargo, para los metanálisis de diferencias de medias estandarizadas, la selección de una desviación estándar extremadamente grande sesgará el resultado hacia una falta de efecto. Hay disponibles alternativas más complicadas para utilizar múltiples posibles desviaciones estándar. Por ejemplo, Marinho y cols. implementaron una regresión lineal del logaritmo (desviación estándar) sobre el logaritmo (media) debido a la fuerte relación lineal entre ellos (Marinho 2003).

Todas las técnicas de imputación incluyen hacer suposiciones acerca de estadísticas desconocidas y es preferible evitar su uso siempre que sea posible. Si la mayoría de los estudios de un metanálisis tiene desviaciones estándar faltantes, estos valores no se deberían imputar. Sin embargo, la imputación puede ser razonable para una pequeña proporción de estudios que incluyan una pequeña proporción de datos, si la misma permite combinarlos con otros estudios para los cuales hay datos completos disponibles. Se deben utilizar análisis de sensibilidad para evaluar la repercusión del cambio de las suposiciones hechas.

16.1.3.2 Imputación de las desviaciones estándar para los cambios a partir del valor inicial

Un caso especial de desviaciones estándar faltantes es para los cambios a partir del valor inicial. A menudo sólo está disponible la siguiente información:

	Valor inicial	Final	Cambio
Intervención experimental (tamaño de muestra)	media, DE	media, DE	media
Intervención control (tamaño de muestra)	media, DE	media, DE	media

Es de señalar que el cambio medio en cada grupo siempre se puede obtener al sustraer la media final de la media inicial, incluso si la misma no se presenta de forma explícita. Sin embargo, la información en esta tabla *no* permite calcular la desviación estándar de los cambios. No es posible conocer si los cambios fueron muy consistentes o muy variables. Alguna otra información en un artículo puede ayudar a determinar la desviación estándar de los cambios. Si se presentan los análisis estadísticos que comparan los cambios (p.ej. los intervalos de confianza, los errores estándar, los valores de t, los valores de p, los valores de f) se pueden utilizar las técnicas descritas en el Capítulo 7 (Sección 7.7.3).

Cuando no hay información disponible suficiente para calcular las desviaciones estándar para los cambios, éstas se pueden imputar. Cuando las desviaciones estándar de los cambios a partir del valor inicial para la misma medida de resultado están disponibles de otros estudios en la revisión, puede ser razonable utilizarlas en lugar de las desviaciones estándar faltantes. Sin embargo, lo apropiado de utilizar una desviación estándar de otro estudio se basa en si los estudios utilizaron la misma escala de medición, tuvieron el mismo grado de error de

medición y tuvieron los mismos períodos de tiempo (entre la medición de los valores iniciales y finales).

Es posible utilizar la siguiente técnica alternativa para imputar las desviaciones estándar faltantes para los cambios a partir de los valores iniciales (Follmann 1992, Abrams 2005). Un número habitualmente no publicado, conocido como coeficiente de correlación, describe cuán similares fueron las mediciones iniciales y finales entre los participantes. Aquí se describe (1) cómo calcular el coeficiente de correlación de un estudio que se describe con considerable detalle y (2) cómo imputar una desviación estándar del cambio a partir del valor inicial en otro estudio, al utilizar un coeficiente de correlación imputado. Es de señalar que los métodos en (2) son aplicables a los coeficientes de correlación obtenidos mediante (1) y a los coeficientes de correlación obtenidos por otras vías (por ejemplo, mediante argumentos razonados). Estos métodos se deberían utilizar de forma limitada porque nunca es posible asegurar que una correlación imputada sea apropiada (las correlaciones entre los valores iniciales y finales, por ejemplo, disminuyen cuando aumenta el tiempo entre las mediciones iniciales y finales, y dependen de los resultados y las características de los participantes). Una alternativa a estos métodos es sencillamente utilizar una comparación de las mediciones finales, la cual en un ensayo aleatorizado en teoría calcula la misma cantidad que la comparación de los cambios a partir del valor inicial.

(1) Cálculo de un coeficiente de correlación a partir de un estudio descrito con considerable detalle

Suponga que hay un estudio disponible que presenta las medias y las desviaciones estándar para el cambio, así como para las mediciones iniciales y finales, por ejemplo:

	Inicial	Final	Cambio
Intervención experimental (tamaño de muestra 129)	media = 15,2 DE = 6,4	media = 16,2 DE = 7,1	media = 1,0 DE = 4,5
Intervención control (tamaño de muestra 135)	media = 15,7 DE = 7,0	media = 17,2 DE = 6,9	media = 1,5 DE = 4,2

En este estudio está disponible el análisis del cambio a partir del valor inicial, si se utilizan solamente los datos de la columna final. Sin embargo, se pueden utilizar otros datos del estudio para calcular los dos coeficientes de correlación, uno para cada grupo de intervención. Si se utiliza la siguiente notación:

	Inicial	Final	Cambio
Intervención experimental (tamaño de muestra N_E)	$M_{E,inicial}$, $DE_{E,inicial}$	$M_{E,final}$, $DE_{E,final}$	$M_{E,cambio}$, $DE_{E,cambio}$
Intervención control (tamaño de muestra N_C)	$M_{C,\ inicial}$, $DE_{C,\ inicial}$	$M_{C,final}$, $DE_{C,final}$	$M_{C,cambio}$, $DE_{C,cambio}$

El coeficiente de correlación en el grupo experimental, $Corr_E$, se puede calcular como:

$$Corr_E = \frac{SD^2_{E,baseline} + SD^2_{E,final} - SD^2_{E,change}}{2 \times SD_{E,baseline} \times SD_{E,final}};$$

y de forma similar para la intervención control, para obtener el $Corr_C$. En el ejemplo, esto se convierte en

$$Corr_E = \frac{6,4^2 + 7,1^2 - 4,5^2}{2 \text{ x } 6,4 \text{ x } 7,1} = 0,78,$$

$$Corr_C = \frac{7,0^2 + 6,9^2 - 4,2^2}{2 \text{ x } 7,0 \text{ x } 6,9} = 0,82.$$

Cuando la desviación estándar inicial o final no estén disponibles, una se puede sustituir por la otra, siempre que sea razonable suponer que la intervención no altera la variabilidad de la medida de resultado. Los coeficientes de correlación se encuentran entre -1 y 1. Si se obtiene un valor menor de 0,5 no tiene valor utilizar el cambio a partir del valor inicial y un análisis de los valores finales será más preciso. Si se supone que los coeficientes de correlación de los dos grupos de intervención son similares, un promedio sencillo proporcionará una medida razonable de la semejanza de las mediciones iniciales y finales entre todos los individuos en el estudio (el promedio de 0,78 y 0,82 para el ejemplo es 0,80). Si los coeficientes de correlación difieren entonces los tamaños de muestra son demasiado pequeños para una estimación fiable, la intervención afecta la variabilidad en las medidas de resultado o el efecto de la intervención depende del nivel inicial, y es mejor evitar el uso de la imputación. Antes de realizar la imputación se recomienda que se computen los coeficientes de correlación para muchos (si no todos) los estudios en el metanálisis y se señale si son consistentes o no. La imputación se debería realizar solamente como un análisis muy tentativo cuando las correlaciones son inconsistentes.

(2) Imputación de la desviación estándar del cambio a partir del valor inicial mediante un coeficiente de correlación

Ahora considere un estudio para el cual falta la desviación estándar de los cambios a partir del valor inicial. Cuando se conocen las desviaciones estándar inicial y final, es posible imputar la desviación estándar faltante mediante un valor imputado, Corr, para el coeficiente de correlación. El valor Corr se puede imputar de otro estudio en el metanálisis (mediante el método [1] anterior), se puede imputar de algún otro lugar o se puede hipotetizar según un argumento razonado. En todas estas situaciones se debería realizar un análisis de sensibilidad, en el que se prueben diferentes valores de Corr, para determinar si el resultado general del análisis es sólido para el uso de los coeficientes de correlación imputados.

Para imputar una desviación estándar de los cambios a partir del valor inicial para la intervención experimental, use

$$DE_{E,cambio} = \sqrt{DE^2_{E,inicial} + DE^2_{E,final} - (2\text{x Corr x } DE_{E,inicial} \text{ x } DE_{E,final})},$$

y de manera similar para la intervención control. Nuevamente, si alguna de las desviaciones estándar (inicial y final) no está disponible, una se pudiera sustituir por la otra si fuera razonable suponer que la intervención no altera la variabilidad de la medida de resultado.

Como ejemplo, a partir de los siguientes datos:

	Inicial	Final	Cambio
Intervención experimental (tamaño de muestra 35)	Media =12,4 DE= 4,2	media = 15,2 DE = 3,8	media = 2,8
Intervención control (tamaño de muestra 38)	media = 10,7 DE = 4,0	media = 13,8 DE = 4,4	media = 3,1

Mediante un coeficiente de correlación imputado de 0,80, es posible imputar la desviación estándar del cambio a partir del valor inicial en el grupo control como:

$$DE_{C,cambio} = \sqrt{4,0^2 + 4,4^2 - (2 \text{ x } 0,80 \text{ x } 4,0 \text{ x } 4,4)} \quad 2,68.$$

16.2 Aspectos del análisis por intención de tratar

16.2.1 Introducción

A menudo algunos participantes se excluyen de los análisis de los ensayos aleatorizados, debido a que se perdieron durante el seguimiento y no se obtuvieron resultados, o debido a que hubo alguna desviación del protocolo, como que recibieron el tratamiento equivocado (o ningún tratamiento), o hubo falta de cumplimiento o falta de elegibilidad. De manera alternativa, puede ser imposible medir ciertos resultados para todos los participantes debido a que su disponibilidad depende de otro resultado (ver Sección 16.2.4). Como se discute en el Capítulo 8 (Sección 8.12), la estimación del efecto de una intervención puede estar sesgada si se excluyen del análisis algunos pacientes asignados al azar. El análisis por intención de tratar tiene como objetivo incluir todos los participantes asignados al azar en un ensayo, independientemente de lo que sucedió posteriormente (Newell 1992, Lewis 1993). Generalmente se prefieren los análisis por intención de tratar porque no están sesgados y porque abordan preguntas más pragmáticas y relevantes clínicamente.

En el Capítulo 8 (Sección 8.12) se describen los siguientes principios de los análisis por intención de tratar:

1. Mantener a los participantes en los grupos de intervención a los cuales se asignaron al azar, independientemente de la intervención que realmente recibieron.
2. Medir datos de resultado en todos los participantes.
3. Incluir en el análisis a todos los participantes asignados al azar.

No hay un consenso claro acerca de si se deberían aplicar todos los criterios (Hollis 1999). Aunque el primero se acepta ampliamente, a menudo el segundo es imposible y el tercero es conflictivo, ya que incluir los participantes cuyos resultados se desconocen (principalmente debido a las pérdidas durante el seguimiento) incluye la imputación (“relleno”) de los datos faltantes (ver Sección 16.1.2).

A menudo un análisis en el cual los datos se analizan para cada uno de los participantes de los cuales se obtuvieron resultados se describe como un análisis de casos disponibles. Algunos ensayos publicados presentan análisis de los resultados solamente de los participantes que completaron el ensayo *y* que cumplieron con la intervención asignada (o recibieron cierta parte de la misma). Algunos autores llaman incorrectamente a este análisis un análisis por intención de tratar, pero de hecho es un análisis por protocolo. Además, algunos autores analizan a los participantes solamente según las intervenciones reales recibidas, independientemente de las asignaciones al azar (análisis por tratamiento recibido). En general no es apropiado aceptar lo que el autor describe en su estudio como análisis por intención de tratar; esta valoración se debería basar en la información detallada proporcionada.

Muchas personas (pero no todas) consideran que los análisis de casos disponibles y por intención de tratar no son apropiados cuando se evalúan efectos no planificados (adversos), ya que es incorrecto atribuir los mismos a un tratamiento que alguien no recibió. Como los análisis por intención de tratar tienden a sesgar los resultados hacia la falta de diferencia, es posible que no sean los más apropiados cuando se intenta establecer la equivalencia o la no inferioridad de un tratamiento.

En la mayoría de las situaciones los revisores deberían intentar extraer de los artículos los datos que permitan al menos un análisis de casos disponibles. De ser posible se deberían "reincluir" las exclusiones evitables. En algunas situaciones poco frecuentes es posible crear un verdadero análisis por intención de tratar a partir de la información presentada en el texto y las tablas del artículo, o mediante la obtención de información adicional del autor acerca de los participantes que se siguieron pero se excluyeron del ensayo. Si es posible hacerlo sin imputar los resultados del estudio, se debería hacer.

Por otra parte, pudiera parecer que el análisis por intención de tratar se puede producir mediante la imputación. Esto incluye hacer suposiciones acerca de los resultados de los participantes de los cuales no se registraron resultados. Sin embargo, muchos análisis de imputación difieren de los análisis de casos disponibles sólo en que tiene una inflación no justificada en la precisión aparente. La evaluación de los resultados de los estudios en presencia de más que los mínimos datos faltantes es finalmente un problema de sentido común, como se discute en el Capítulo 8 (Sección 8.12). El análisis estadístico no puede compensar de forma fiable los datos faltantes (Unnebrink 2001). Es probable que ninguna suposición refleje adecuadamente la realidad, y la repercusión de cualquier suposición se debería evaluar al probar más de un método como un análisis de sensibilidad (ver Capítulo 9, Sección 9.7).

En las dos secciones siguientes se consideran algunas formas de tener en cuenta las observaciones faltantes para los resultados dicotómicos o continuos. Aunque es posible la imputación, hasta el momento una decisión sensible en la mayoría de los casos es incluir solamente datos para los participantes cuyos resultados se conocen, y abordar la repercusión potencial de los datos faltantes en la evaluación del riesgo de sesgo (Capítulo 8, Sección 8.12). Cuando se utilice la imputación, los métodos y las suposiciones para la imputación de los datos de los abandonos se deberían describir en la sección "Métodos" del protocolo y la revisión.

Si hay datos disponibles de los participantes individuales, se pueden considerar análisis de sensibilidad detallados. Los revisores en este caso se pueden referir a la extensa bibliografía sobre el procesamiento de los datos faltantes en los ensayos clínicos (Little 2004). En general los participantes excluidos de los análisis en las publicaciones se deberían reincluir cuando sea posible, como en el caso en el que hay datos de participantes individuales disponibles

(Stewart 1995). Se les debería solicitar información a los autores del ensayo cuando no haya suficientes detalles disponibles en los ensayos publicados para reincluir los participantes excluidos en los análisis.

16.2.2 Aspectos del análisis por intención de tratar para los datos dicotómicos

Las proporciones de participantes para los cuales no se obtuvieron datos de resultado siempre se deberían obtener e informar en una tabla "Riesgo de sesgo"; es de señalar que las proporciones pueden variar según el resultado y el grupo asignado al azar. Sin embargo, no hay consenso sobre la mejor forma de tratar a estos participantes en un análisis. Hay dos opciones básicas, y se debería utilizar una opción factible como análisis principal y como base para el análisis de sensibilidad (ver más adelante y Capítulo 9, Sección 9.7).

- Análisis de casos disponibles: Incluye solamente datos de los resultados conocidos, y utiliza como denominador el número total de personas que tiene datos registrados para el resultado particular. La variación en el grado de datos faltantes entre los estudios se puede considerar como una fuente potencial de heterogeneidad.
- Análisis por intención de tratar mediante imputación: Basa el análisis en el número total de participantes asignados al azar independientemente de cómo los autores originales del estudio analizaron los datos. Esto incluirá la imputación de resultados para los participantes faltantes. Hay numerosos enfoques para imputar los datos de resultados dicotómicos. Un enfoque frecuente es suponer que todos los participantes faltantes presentaron el evento o que todos los participantes faltantes no presentaron el evento. Un enfoque alternativo es imputar los datos según la tasa de evento observada en el grupo control, o según las tasas de eventos entre los que completaron en los grupos separados (éste último proporciona la misma estimación del efecto de la intervención pero los resultados presentan una inflación no justificada de la precisión de las estimaciones del efecto). La decisión entre estas suposiciones se debería basar en la valoración clínica. A los estudios con resultados imputados se les puede dar una mayor ponderación que la que merecen si se introdujeron como datos dicotómicos en RevMan. Es posible determinar ponderaciones más apropiadas (Higgins 2008); se recomienda la consulta con un estadístico. Sin embargo, ninguna de estas suposiciones es probable que refleje la realidad, excepto la imputación de "fracasos" en algunos ámbitos como los ensayos de abandono del hábito de fumar, por lo que en general no se recomienda el enfoque de la imputación.

La repercusión potencial de los datos faltantes sobre los resultados se debería considerar cuando se interpreten los resultados de la revisión. Esto dependerá del grado de "ausencia", la frecuencia de los eventos y el tamaño de la estimación combinada del efecto. Gamble y Hollis recomiendan un análisis de sensibilidad para los resultados dicotómicos basado en la consideración de los escenarios del "mejor caso" y el "peor caso" (Gamble 2005). El escenario del "mejor caso" es que todos los participantes con resultados faltantes del grupo de intervención experimental tuvieron resultados favorables, y que todos los participantes con resultados faltantes del grupo de intervención control tuvieron resultados deficientes; el escenario del "peor caso" es lo contrario. El análisis de sensibilidad disminuye la ponderación de los estudios en los cuales la discrepancia entre los escenarios del "mejor caso" y el "peor caso" es alta, aunque esta disminución de la ponderación puede ser demasiado exagerada.

Un análisis de sensibilidad mas factible considera explícitamente cuáles pueden haber sido las tasas de evento en los datos faltantes. Por ejemplo, suponga que se utilizó un análisis de casos disponibles y un estudio particular tiene un riesgo del 20% en el brazo intervención y un riesgo del 15% en el brazo control. Un análisis de casos disponibles supone implícitamente que las mismas fracciones se aplican a los datos faltantes, por lo que tres análisis de sensibilidad adecuados para comparar con este análisis pueden considerar que el riesgo en los

datos faltantes es del 15% en ambos brazos, o del 15% y el 10% en los brazos control y experimental respectivamente, o del 20% y del 10% respectivamente. De manera alternativa, suponga que en el análisis principal todos los valores faltantes se han imputado como eventos. Un análisis de sensibilidad para comparar con este análisis pudiera considerar el caso de que, por ejemplo, el 10% de los participantes faltantes presentaron el evento, o el 10% en el brazo de intervención y el 5% en el brazo control. Se han considerado los enfoques gráficos a los análisis de sensibilidad (Hollis 2002).

Higgins y cols. recomiendan un enfoque alternativo que puede incorporar motivos específicos para los datos faltantes, el cual considera factible riesgos de eventos entre los participantes faltantes en relación con los riesgos entre los observados (Higgins 2008). White y cols. (White 2008a, White 2008b) consideran los enfoques bayesianos que automáticamente disminuyen la ponderación de los estudios con más datos faltantes.

16.2.3 Aspectos del análisis por intención de tratar para los datos continuos

En los análisis completos por intención de tratar todos los participantes que no recibieron una intervención según el protocolo, así como los que se perdieron del seguimiento, se incluyen en el análisis. La inclusión de éstos en un análisis requiere que estén disponibles las medias y las desviaciones estándar del resultado para todos los participantes asignados al azar. Al igual que para los datos dicotómicos, siempre se deberían obtener e informar las tasas de abandono en una tabla de “Riesgo de sesgo”. Nuevamente, hay dos opciones básicas y en cualquier caso se debería realizar un análisis de sensibilidad (ver Capítulo 9, Sección 9.7):

- Análisis de casos disponibles: Incluye solamente datos de los resultados conocidos. La repercusión potencial de los datos faltantes sobre los resultados se debería considerar cuando se interpreten los resultados de la revisión. Esto dependerá del grado de “ausencia”, la estimación combinada del efecto del tratamiento y de la variabilidad de los resultados. La variación en el grado de datos faltantes también se puede considerar una fuente potencial de heterogeneidad.
- Análisis por intención de tratar mediante imputación: Basa el análisis en el número total de participantes asignados al azar independientemente de cómo los autores originales del estudio analizaron los datos. Esto incluirá la imputación de resultados para los participantes faltantes. Los enfoques para imputar los datos continuos faltantes en el contexto de un metanálisis han recibido poca atención en la bibliografía metodológica. En algunas situaciones puede ser posible utilizar los enfoques estándar (aunque cuestionables) de la “última observación realizada” o, para los resultados de cambio a partir del valor inicial, suponer que no se produjo un cambio, pero tales enfoques generalmente requieren el acceso a los datos crudos de los participantes. No se recomienda la inflación del tamaño de muestra de los datos disponibles hasta el número total de participantes asignados al azar porque este procedimiento inflará artificialmente la precisión de la estimación del efecto.

Una forma sencilla de realizar un análisis de sensibilidad para los datos continuos es suponer una diferencia fija entre la media real para los datos faltantes y la media supuesta por el análisis. Por ejemplo, después de un análisis de casos disponibles, se pudiera considerar cómo los resultados habrían diferido si los datos faltantes en el brazo intervención tuvieran como promedio dos unidades *más* que los datos observados en el brazo intervención, y si los datos faltantes en el brazo control hubieran tenido como promedio 2 unidades *menos* que los datos observados en el brazo control. Se ha considerado un enfoque bayesiano que disminuye automáticamente la ponderación de los estudios con más datos faltantes (White 2007).

16.2.4 Resultados condicionales disponibles solamente para subgrupos de participantes

Algunos resultados de los estudios pueden ser aplicables solamente a una proporción de los participantes. Por ejemplo, en los ensayos de subfertilidad a menudo se informa la proporción de mujeres con embarazos clínicos que presentan un aborto espontáneo después del tratamiento. Por definición este resultado excluye a las participantes que no alcanzaron un estado intermedio (embarazo clínico), por lo que la comparación no se realiza para todas las participantes asignadas al azar. Como regla general es mejor redefinir estos resultados de manera que el análisis incluya a todos los participantes asignados al azar. En este ejemplo, el resultado sería si la mujer tuvo un "embarazo exitoso" (logró el embarazo y alcanzó, por ejemplo, las 24 semanas o el término). Otro ejemplo lo proporciona un resultado de morbilidad medido a medio o largo plazo (p.ej. el desarrollo de enfermedad pulmonar crónica), cuando hay una posibilidad inequívoca de muerte que impide la evaluación de la morbilidad. Una manera conveniente de tratar estas situaciones es combinar los resultados, por ejemplo "muerte o enfermedad pulmonar crónica).

Algunos problemas difíciles de tratar surgen cuando un resultado continuo (como una medida de la capacidad funcional o la calidad de vida después de un accidente cerebrovascular) se mide sólo en los que sobreviven al final del seguimiento. Existen dos alternativas no satisfactorias: (a) imputar puntuaciones cero de capacidad funcional para los que murieron (lo que puede que no represente apropiadamente el estado de muerte y hará que el resultado esté intensamente sesgado), y (b) analizar los datos disponibles (que se pueden interpretar como una comparación no aleatorizada aplicable solamente a los supervivientes). Los resultados del análisis se deberían interpretar teniendo en cuenta cualquier disparidad en la proporción de muertes entre los dos grupos de intervención.

16.3 Ensayos aleatorizados por grupo

16.3.1 Introducción

En los ensayos aleatorizados por grupo, grupos de individuos en lugar de individuos se asignan al azar a diferentes intervenciones. Los ensayos aleatorizados por grupos también se conocen como ensayos aleatorizados grupales. Se dice que la "unidad de asignación" es el grupo. Los grupos pueden ser por ejemplo escuelas, pueblos, centros de atención primaria o familias. Dichos ensayos se pueden realizar por uno de varios motivos. Puede ser para evaluar el efecto grupal de una intervención, por ejemplo la inmunidad grupal de una vacuna. Se puede realizar para evitar "contaminación" entre las intervenciones cuando los participantes de los ensayos se tratan dentro del mismo ámbito; por ejemplo, en un ensayo que evalúa una intervención dietética, se pueden asignar al azar familias en lugar de individuos. Se puede utilizar un diseño aleatorizado por grupo sencillamente por conveniencia.

Una de las consecuencias principales de un diseño por grupo es que los participantes dentro de cualquier grupo a menudo tienden a responder de una manera similar y por lo tanto se deja de suponer que sus datos son independientes entre sí. Sin embargo, muchos de estos estudios se analizan de manera incorrecta ya que se considera que la unidad de asignación han sido los participantes individuales. A esto se le llama a menudo "error de unidad de análisis" (Whiting-O'Keefe 1984) porque la unidad de análisis es diferente de la unidad de asignación. Si se ignora el agrupamiento y los ensayos por grupo se analizan como si se hubieran asignado al azar individuos, los valores de p resultantes serán artificialmente pequeños. Esto puede dar lugar a conclusiones falsas positivas de que la intervención tuvo un efecto. En el contexto de un metanálisis los estudios en los cuales se ha ignorado el

agrupamiento tendrán intervalos de confianza extremadamente estrechos y recibirán más ponderación que la apropiada en un metanálisis. Esta situación también puede surgir si los participantes se asignan a intervenciones que luego se aplican a partes de los sujetos (por ejemplo, a ambos ojos o a varios dientes), o si se hacen observaciones repetidas en un participante. Si el análisis se realiza por unidades individuales (por ejemplo, cada diente o cada observación) sin tener en cuenta que los datos se agruparon dentro de los participantes, entonces puede ocurrir un error de unidad de análisis.

Hay varias fuentes útiles de información sobre los ensayos aleatorizados por grupo (Murray 1995, Donner 2000). Está disponible una discusión detallada de la incorporación de ensayos aleatorizados por grupo en un metanálisis (Donner 2002), la cual es una discusión más técnica del problema (Donner 2001). White y Thomas (White 2005) señalan consideraciones especiales para el análisis de las diferencias de medias estandarizadas de los ensayos aleatorizados por grupo.

16.3.2 Evaluación del riesgo de sesgo en los ensayos aleatorizados por grupo

En los ensayos aleatorizados por grupo los sesgos particulares a considerar son: (i) sesgo de reclutamiento; (ii) desequilibrio basal; (iii) pérdida de los grupos; (iv) análisis incorrecto; y (v) comparabilidad con ensayos con asignación al azar individual.

(i) El sesgo de reclutamiento puede ocurrir cuando los individuos se reclutan en el ensayo después que los grupos se asignaron al azar, ya que el conocimiento de si cada grupo es un grupo “intervención” o “control” pudiera afectar el tipo de participantes reclutado. Farrin y cols. mostraron un reclutamiento diferencial de los participantes en un ensayo de dolor lumbar con asignación al azar según el centro de atención primaria; un mayor número de participantes menos graves se reclutaron en las prácticas de “tratamiento activo” (Farrin 2005). Puffer y cols revisaron 36 ensayos aleatorizados por grupo y encontraron un posible sesgo de reclutamiento en 14 (39%) (Puffer 2003).

(ii) A menudo los ensayos aleatorizados por grupo asignan todos los grupos a la misma vez, por lo que habitualmente la falta de ocultación de una secuencia de asignación no es un problema. Sin embargo, debido a que se asignan al azar pequeños números de grupos, existe la posibilidad de desequilibrios basales por azar entre los grupos asignados al azar, en términos de los grupos o los individuos. Aunque no es una forma de sesgo como tal, el riesgo de diferencias basales se puede reducir mediante el uso de asignación de los grupos estratificada o pareada. Describir la comparabilidad inicial de los grupos o el ajuste estadístico según las características basales puede ayudar a reducir las preocupaciones acerca de los efectos del desequilibrio basal.

(iii) Ocasionalmente se pierden en un ensayo grupos completos, y los mismos se deberían omitir del análisis. Al igual que los datos de resultado faltantes en los ensayos con asignación al azar individual, este hecho puede producir sesgo. Además, los resultados faltantes para individuos dentro de los grupos también pueden dar lugar a riesgo de sesgo en los ensayos aleatorizados por grupo.

(iv) Muchos ensayos aleatorizados por grupo se analizan mediante métodos estadísticos incorrectos y no tienen en cuenta el agrupamiento. Por ejemplo, Eldridge y cols. revisaron 152 ensayos aleatorizados por grupo en la atención primaria, de los cuales el 41% no tuvo en cuenta el agrupamiento en sus análisis (Eldridge 2004). Dichos análisis crean un “error de unidad de análisis” y producen resultados demasiado precisos (el error estándar de la estimación del efecto de la intervención es demasiado pequeño) y los valores de p son

demasiado pequeños. Estos no provocan estimaciones sesgadas del efecto pero, si no se corrigen, recibirán demasiada ponderación en un metanálisis. En la Sección 16.3.6 se indican métodos aproximados para corregir los resultados de ensayos que no permiten el agrupamiento. Los revisores pueden implementar algunos de ellos.

(v) En un metanálisis que incluya ensayos aleatorizados por grupo e individuales o que incluya ensayos aleatorizados por grupo con diferentes tipos de grupos, es necesario considerar las posibles diferencias entre los efectos de la intervención que se calculan. Por ejemplo, en un ensayo de una vacuna de enfermedades infecciosas, es de esperar que una vacuna aplicada a todos los individuos en una comunidad sea más efectiva que si la vacuna se aplica solamente a la mitad de las personas. Hahn y cols proporcionan otro ejemplo y analizaron una revisión Cochrane sobre protectores de cadera (Hahn 2005). Los ensayos por grupo mostraron grandes efectos positivos, mientras que los ensayos con asignación al azar individual no mostraron un beneficio claro. Una posibilidad es que hubo un “efecto de grupo” en los ensayos aleatorizados por grupo (que a menudo se realizaron en residencias de ancianos, donde puede haber mejorado el cumplimiento con el uso de los protectores). En general dicha “contaminación” daría lugar a una subestimación del efecto. Por lo tanto, si el efecto de una intervención se demuestra incluso a pesar de la contaminación en los ensayos que no tuvieron una asignación al azar por grupo, se puede establecer una conclusión fiable acerca de la presencia de un efecto. Sin embargo, es probable que se subestime el tamaño del efecto. La contaminación y los “efectos de grupo” pueden ser diferentes para diferentes tipos de grupos.

16.3.3 Métodos de análisis para los ensayos aleatorizados por grupo

Una manera de evitar los errores de la unidad de análisis en los ensayos aleatorizados por grupo es realizar el análisis al mismo nivel de la asignación, mediante una medición resumida de cada grupo. Entonces el tamaño de muestra es el número de grupos y el análisis se realiza como si el ensayo tuviera una asignación al azar individual (aunque los grupos se convierten en individuos). Sin embargo, esto pudiera reducir considerable e innecesariamente el poder estadístico del estudio, según el número y el tamaño de los grupos.

Por otra parte, actualmente existen métodos estadísticos que permiten el análisis a nivel del individuo a la vez que tienen en cuenta el agrupamiento en los datos. La información ideal a extraer de un ensayo aleatorizado por grupo es una estimación directa de la medida de efecto requerida (por ejemplo, un *odds ratio* con su intervalo de confianza) de un análisis que tenga en cuenta adecuadamente el diseño por grupo. Dicho análisis se pudiera basar en un “modelo multinivel”, un “análisis de componentes de la varianza” o puede utilizar “ecuaciones de estimación generalizada” (en inglés, GEE), entre otras técnicas. Se recomienda buscar asesoría estadística para determinar si el método utilizado es apropiado. Las estimaciones del efecto y sus errores estándar a partir de los análisis correctos de los ensayos aleatorizados por grupo se pueden explorar en un metanálisis mediante el método de la varianza inversa genérica en RevMan.

16.3.4 Análisis aproximados de los ensayos aleatorizados por grupo para un metanálisis: tamaños de muestra efectivos

Desafortunadamente muchos ensayos aleatorizados por grupos no lograron en el pasado proporcionar análisis apropiados. Estos se analizan frecuentemente como si la asignación al azar se hubiera realizado en individuos en lugar de grupos. Si esta es la situación, se podrían realizar análisis aproximadamente correctos si es posible extraer la siguiente información:

- el número de grupos (o grupos) asignados a cada grupo de intervención; o el tamaño promedio (media) de cada grupo;
- los datos de resultado que ignoran el diseño por grupo para el número total de individuos (por ejemplo, el número o la proporción de individuos con eventos, o las medias y las desviaciones estándar);
- una estimación del coeficiente de correlación intragrupo (o intraclase) (CCI).

El CCI es una estimación de la variabilidad relativa dentro y entre los grupos (Donner 1980). Describe la "semejanza" de los individuos dentro del mismo grupo. De hecho, pocas veces está disponible en las publicaciones. Un enfoque frecuente es utilizar estimaciones externas obtenidas de estudios similares, y hay numerosos recursos disponibles que proporcionan ejemplos de CCI (Ukoumunne 1999, Campbell 2000, Health Services Research Unit 2004). Los CCI pueden parecer pequeños comparados con otros tipos de correlaciones: son típicos los valores menores de 0,05. Sin embargo, incluso valores pequeños pueden tener una repercusión significativa sobre la amplitud de los intervalos de confianza (y por lo tanto, sobre las ponderaciones en un metanálisis), particularmente si el tamaño del grupo es grande. Las investigaciones empíricas han observado que tamaños más grandes de los grupos se asocian con CCI más pequeños (Ukoumunne 1999).

Un análisis aproximadamente correcto es el siguiente: la idea es reducir el tamaño de cada ensayo a su "tamaño de muestra efectivo" (Rao 1992). El tamaño de muestra efectivo de un grupo de intervención único en un ensayo aleatorizado por grupo es su tamaño de muestra original dividido por una cantidad llamada "efecto del diseño". El efecto del diseño es

$$1 + (M - 1)\ CCI,$$

donde M es el tamaño promedio del grupo y CCI es el coeficiente de correlación intragrupo. Habitualmente se supone un efecto del diseño común entre los grupos de intervención. Para los datos dicotómicos se deberían dividir el número de participantes y el número que presenta el evento por el mismo efecto del diseño. Como los datos resultantes se tienen que redondear a números enteros para introducirlos en RevMan, este enfoque puede ser inadecuado para ensayos pequeños. Para los datos continuos sólo es necesario reducir el tamaño de muestra; las medias y las desviaciones estándar no deberían cambiarse.

16.3.5 Ejemplo de incorporación de un ensayo aleatorizado por grupo

Como ejemplo, considere un ensayo aleatorizado por grupo que asignó al azar diez aulas de escuelas con 295 niños a un grupo de intervención y 11 aulas con 330 niños a un grupo control. El número de resultados favorables entre los niños, si se ignora el agrupamiento, es

Intervención: 63/295

Control: 84/330.

Imagine que se obtuvo un coeficiente de correlación intragrupo de 0,02 de una fuente externa fiable. El tamaño promedio del grupo en el ensayo es (295+330)/(10+11) = 29,8. El efecto del diseño para el ensayo como un todo es entonces 1 + (M – 1) CCI = 1 + (29,8 – 1)×0,02 = 1,576. El tamaño de muestra efectivo en el grupo de intervención es 295 / 1,576 = 187,2 y para el grupo control es 330 / 1,576 = 209,4.

Al aplicar el efecto del diseño al número de eventos se producen los siguientes resultados:

Intervención: 40,0/187,2

Control: 53,3/209,4.

Una vez que los ensayos se han reducido a su tamaño de muestra efectivo los datos se pueden introducir en RevMan como, por ejemplo, resultados dicotómicos o resultados continuos. Los resultados del ensayo ejemplo se pueden introducir como

Intervención: 40/187

Control: 53/209.

16.3.6 Análisis aproximados de los ensayos aleatorizados por grupo para un metanálisis: inflación de los errores estándar

Una clara desventaja del método descrito en la Sección 16.3.4 es la necesidad de redondear los tamaños de muestra efectivos a números enteros. Un enfoque algo más flexible, equivalente a calcular los tamaños de muestra efectivos, es multiplicar los errores estándar de la estimación del efecto (de un análisis que ignore el agrupamiento) por la raíz cuadrada del efecto del diseño. El error estándar se puede calcular de un intervalo de confianza (ver Capítulo 7, Sección 7.7.7). Se pueden utilizar los análisis estándar de los resultados dicotómicos o de los resultados continuos para obtener estos intervalos de confianza mediante RevMan. El metanálisis que utiliza las varianzas infladas se puede realizar con RevMan y el método de la varianza inversa genérica.

Como ejemplo, el *odds ratio* (OR) de un estudio con los resultados

Intervención: 63/295

Control: 84/330

es OR = 0,795 (IC del 95%: 0,548 a 1,154). Mediante los métodos descritos en el Capítulo 7 (Sección 7.7.7.3) es posible determinar a partir de estos resultados que el logaritmo *odds ratio* es lnOR = –0,23 con un error estándar 0,19. Al utilizar el mismo efecto del diseño 1,576 como en la Sección 16.3.5, un error estándar inflado que toma en cuenta el agrupamiento está dado por 0,19 × √1,576 = 0,24. El log *odds ratio* (–0,23) y su error estándar inflado (0,24) se pueden introducir en RevMan como una variable genérica del inverso de la varianza.

16.3.7 Aspectos de la incorporación de ensayos aleatorizados por grupo

Los ensayos aleatorizados por grupo pueden, en principio, combinarse con ensayos aleatorizados individuales en el mismo metanálisis. Se debería tener en cuenta la posibilidad de diferencias importantes en los efectos que se evalúan entre los diferentes tipos de ensayos. A menudo existen buenos motivos para realizar ensayos aleatorizados por grupo y los mismos se deberían examinar. Por ejemplo, en el tratamiento de las enfermedades infecciosas una intervención aplicada a todos los individuos en una comunidad puede ser más efectiva que el tratamiento aplicado a individuos seleccionados (asignados al azar) dentro de la comunidad porque la misma puede reducir la posibilidad de reinfección.

Los revisores deberían identificar siempre cualquier ensayo aleatorizado por grupo en una revisión y señalar explícitamente cómo procesaron los datos. Deberían realizar análisis de sensibilidad para investigar la solidez de sus conclusiones, especialmente cuando los CCI se han tomado de fuentes externas (ver Capítulo 9, Sección 9.7). Se recomienda buscar apoyo estadístico.

16.3.8 Ensayos aleatorizados individuales con agrupamiento

También pueden surgir problemas relacionados con el agrupamiento en los ensayos aleatorizados individuales. Esto puede suceder cuando el mismo profesional de la salud administra la intervención a un número de participantes en el grupo de intervención. Lee y Thompson analizan este tipo de agrupamiento, que da lugar a situaciones similares a las de los ensayos aleatorizados por grupo (lee 2005a).

16.4 Ensayos cruzados

16.4.1 Introducción

Los ensayos de grupos paralelos asignan cada participante a una intervención única para la comparación con una o más intervenciones alternativas. Por el contrario, los ensayos cruzados asignan cada participante a una secuencia de intervenciones. Un diseño aleatorizado cruzado simple es un diseño “AB/BA” en el cual los participantes se asignan de manera aleatoria inicialmente a la intervención A o a la intervención B, y luego se “cruzan” a la intervención B o la intervención A, respectivamente. Se puede observar que los datos del primer período de un ensayo cruzado representan un ensayo de grupo paralelo, una característica referida en la Sección 16.4.5. Para mantener la uniformidad con el resto del *Manual*, se utilizarán E y C para referirse a las intervenciones, en lugar de A y B.

Los diseños cruzados ofrecen varias ventajas posibles con respecto a los ensayos de grupos paralelos. Entre ellas están (i) que cada participante actúa como su propio control, lo que elimina la variación entre participantes; (ii) que, por lo tanto, se requieren menos participantes para obtener el mismo poder estadístico; y (iii) que cada participante recibe cada intervención, lo que permite determinar la mejor intervención o la preferencia para un participante individual. Senn (Senn 2002) proporciona una introducción fácil de leer sobre los ensayos cruzados. Elbourne y cols. (Elbourne 2002) brindan una discusión más detallada del metanálisis que incluye los ensayos cruzados y Lathyris y cols. (Lathyris 2007) proporcionan algunas pruebas empíricas sobre su inclusión en las revisiones sistemáticas.

16.4.2 Evaluación de la conveniencia de los ensayos cruzados

Los ensayos cruzados son adecuados para evaluar las intervenciones con un efecto temporal en el tratamiento de afecciones crónicas estables. Se utilizan por ejemplo en el estudio de intervenciones para aliviar el asma o la epilepsia. No son apropiados cuando una intervención puede tener un efecto duradero que comprometa el ingreso a los períodos posteriores del ensayo, o cuando una enfermedad tiene una evolución rápida. Las ventajas de los ensayos cruzados se deberían equilibrar con sus desventajas. El problema principal asociado con los ensayos cruzados es el efecto de arrastre (un tipo de interacción de período por intervención). El efecto de arrastre es la situación en la cual los efectos de una intervención administrada en un período persisten en un período posterior, por lo que interfieren con los efectos de una intervención diferente posterior. Muchos ensayos cruzados incluyen un período entre las intervenciones conocido como período de lavado como una forma de reducir el efecto de arrastre. Si un resultado primario es irreversible (por ejemplo, mortalidad o embarazo en un estudio de subfertilidad), generalmente un estudio cruzado no se considera apropiado. Otro problema con los ensayos cruzados es el riesgo de abandono debido a su mayor duración comparados con ensayos de grupos paralelos similares. Las técnicas de análisis para los ensayos cruzados con observaciones faltantes son limitadas. La evaluación del riesgo de sesgo en los ensayos cruzados se discute en la Sección 16.4.3

Al considerar la inclusión de los ensayos cruzados en el metanálisis los autores deberían abordar primero la pregunta de si un ensayo cruzado es un método adecuado para la afección y la intervención en estudio. Por ejemplo, aunque se utilizan frecuentemente en este campo, un grupo de revisores decidió que los ensayos cruzados no eran apropiados para los estudios de la enfermedad de Alzheimer debido a la naturaleza degenerativa de la afección, y sólo incluyeron datos del primer período (Qizilbash 1998). El segundo interrogante a abordar es si existe la probabilidad de un efecto de arrastre importante, lo cual depende en gran medida del sentido común, ya que las técnicas estadísticas para demostrar el efecto de arrastre están lejos de ser satisfactorias. La naturaleza de las intervenciones y la duración de cualquier período de lavado son consideraciones importantes.

Sólo se justifica excluir los ensayos cruzados de una revisión sistemática si el diseño no es apropiado al contexto clínico. Sin embargo, con mucha frecuencia es difícil o imposible extraer datos adecuados de un ensayo cruzado. En la Sección 16.4.5 se esbozan algunas consideraciones e indicaciones para incluir los ensayos cruzados en un metanálisis. Primero se discute cómo se puede ampliar la herramienta “Riesgo de sesgo” descrita en el Capítulo 8 para abordar preguntas específicas a los ensayos cruzados.

16.4.3 Evaluación del sesgo en los ensayos cruzados

Las principales inquietudes acerca del riesgo de sesgo en los ensayos cruzados son: (i) si el diseño cruzado es adecuado; (ii) si existe un efecto de arrastre; (iii) si hay datos disponibles sólo del primer período; (iv) análisis incorrecto; y (v) comparabilidad de los resultados con los de los ensayos de grupos paralelos.

(i) El diseño cruzado es adecuado para estudiar una afección (razonablemente) estable (p.ej. asma) y en la que no se requiere un seguimiento a largo plazo. Por lo tanto, el primer aspecto a considerar es si el diseño cruzado es adecuado para la afección en estudio.

(ii) Una preocupación importante es la posibilidad de un “efecto de arrastre” del tratamiento de un período al siguiente. Un efecto de arrastre significa que la diferencia observada entre los tratamientos depende del orden en el cual se recibieron los mismos; por lo tanto la estimación general del efecto del tratamiento estará afectada (habitualmente subestimada, lo que da lugar a sesgo hacia la nulidad).

Por lo tanto, el uso del diseño cruzado se debería limitar a situaciones en las cuales es poco probable que haya un efecto de arrastre del tratamiento entre los períodos. Sin embargo, es posible que no haya apoyo disponible para esta opción antes de que se realice el ensayo. Los revisores deberían buscar información en los los ensayos acerca de la evaluación del efecto de arrastre. Sin embargo, en una revisión no publicada de 116 ensayos cruzados publicados desde el año 2000 (Mills 2005), el 30% de los estudios discutieron el efecto de arrastre, pero sólo el 12% publicó el análisis.

(iii) En presencia de un efecto de arrastre una estrategia frecuente es basar el análisis solamente en el primer período. Aunque el primer período de un ensayo cruzado es en efecto una comparación de grupos paralelos, el uso de datos de sólo el primer período estará sesgado si, como es probable, la decisión de hacerlo se basa en una prueba de efecto de arrastre. Este “análisis de dos estadios” se ha desacreditado (Freeman 1989) pero aún se utiliza. Además, el uso de solamente el primer período elimina la principal fortaleza del diseño cruzado, la capacidad de comparar tratamientos en individuos.

Se debería considerar que los ensayos cruzados para los cuales sólo están disponibles los datos del primer período tienen riesgo de sesgo, especialmente cuando los investigadores utilizaron explícitamente la estrategia de dos estadios.

(iv) El análisis de un ensayo cruzado debería aprovechar el diseño intra-sujetos, y utilizar alguna forma de análisis pareado (Elbourne 2002). Aunque los autores del ensayo pueden haber analizado los datos pareados, la presentación deficiente puede imposibilitar que los revisores extraigan los datos pareados. Los datos no pareados pueden estar disponibles y generalmente no se relacionan con la estimación del efecto del tratamiento ni con la significación estadística. Por lo tanto, no son una fuente de sesgo, pero en su lugar frecuentemente darán lugar a que el ensayo reciba en un metanálisis una ponderación (mucho) menor que la que merece.

En la revisión mencionada anteriormente (Mills 2005) sólo el 38% de 116 ensayos cruzados realizó un metanálisis de datos pareados.

(v) A falta de efecto de arrastre, los ensayos cruzados deberían estimar el mismo efecto del tratamiento que los ensayos de grupos paralelos. Aunque un estudio describió una diferencia en el efecto del tratamiento encontrada en los ensayos cruzados comparados con los ensayos de grupos paralelos (Khan 1996), los autores habían examinado tratamientos para la infertilidad, un área no apropiada para el diseño cruzado, y un reanálisis cuidadoso no apoyó los hallazgos originales (te Velde 1998).

Otros aspectos a considerar para el riesgo de sesgo en los ensayos cruzados incluyen los siguientes:

- Los participantes pueden abandonar después del primer tratamiento y no recibir el segundo tratamiento. Generalmente estos participantes se retiran del análisis.
- Puede haber diferencias sistemáticas entre los dos períodos del ensayo. Un efecto del período no es demasiado grave, puesto que se aplica por igual a ambos tratamientos, aunque puede indicar que la afección estudiada no es estable.
- Es posible que no esté claro cuántos tratamientos o períodos se utilizaron. Lee no pudo identificar el diseño para 12/64 ensayos cruzados publicados (Lee 2005b).
- No se debería suponer que el orden de los tratamientos fue aleatorizado en un ensayo cruzado. Ocasionalmente se puede encontrar un estudio en el cual está claro que todos los participantes recibieron los tratamientos en el mismo orden. Este ensayo no proporciona una comparación válida de los tratamientos porque puede haber una tendencia de los resultados en el tiempo, además del cambio en los tratamientos.
- La descripción de los abandonos puede ser deficiente, especialmente para los participantes que finalizaron un período de tratamiento. En la revisión de Lee el número de participantes que abandonaron se especificó solamente en nueve de 64 ensayos (Lee 2005b).

Algunas preguntas indicadas para evaluar el riesgo de sesgo en los ensayos cruzados son las siguientes:

- ¿Fue apropiado el uso de un diseño cruzado?
- ¿Está claro que el orden para recibir el tratamiento fue aleatorizado?
- ¿Se puede suponer que el ensayo no estaba sesgado debido al efecto de arrastre?
- ¿Hay datos disponibles no sesgados?

16.4.4 Métodos de análisis para los ensayos cruzados

Si se considera que ni el efecto de arrastre ni los efectos de los períodos son un problema, entonces un análisis apropiado de los datos continuos de un ensayo cruzado de dos períodos y dos intervenciones es una prueba t pareada. Esta evalúa el valor de la "medición en la intervención experimental (E)" menos la "medición en la intervención control (C)" de forma separada para cada participante. La media y el error estándar de estas medidas de diferencias son el fundamento de una estimación del efecto y una prueba estadística. La estimación del efecto se puede incluir en un metanálisis mediante el método de la varianza inversa genérica en RevMan.

Es posible realizar un análisis pareado si están disponibles los datos de cualquiera de los siguientes puntos:

- los datos de participantes individuales de los artículos o por correspondencia con los autores de los ensayos;
- la media y la desviación estándar (o el error estándar) de las diferencias específicas de los participantes entre las mediciones en la intervención experimental (E) y la intervención control (C);
- la diferencia de medias y una de la siguientes: (i) un estadístico t de una prueba t pareada; (ii) un valor de p de una prueba t pareada; (iii) un intervalo de confianza de un análisis pareado;
- un gráfico de mediciones en la intervención experimental (E) y la intervención control (C) de los cuales es posible extraer los valores de los datos individuales, siempre que se puedan identificar como tales las mediciones pareadas para cada individuo.

Para detalles ver Elbourne y cols. (Elbourne 2002).

Si los resultados están disponibles, separados por la secuencia particular que recibió cada participante, los análisis que se ajustan por los efectos de los períodos son sencillos (p.ej. como se esboza en el Capítulo 3 de Senn (Senn 2002).

16.4.5 Métodos para incorporar los ensayos cruzados en un metanálisis

Desafortunadamente la publicación de los ensayos cruzados ha sido muy variable y a menudo no se publican los datos necesarios para incluir un análisis pareado en un metanálisis. Una situación frecuente es que las medias y las desviaciones estándar (o los errores estándar) están disponibles solamente para las mediciones en E y C por separado. Un enfoque sencillo para incorporar los ensayos cruzados en un metanálisis es, por lo tanto, tomar todas las mediciones de los períodos de la intervención E y todas las mediciones de los períodos de la intervención C y analizarlos como si el ensayo fuera de grupos paralelos de E versus C. Este enfoque da lugar a un error de unidad de análisis (ver Capítulo 9, Sección 9.3) y se debería evitar a menos que se pueda demostrar que los resultados se aproximan a los de un análisis pareado, como se describe en la Sección 16.4.4. El motivo para esto es que es probable que los intervalos de confianza sean demasiado amplios y que el ensayo reciba muy poca ponderación, con la posible consecuencia de disfrazar una heterogeneidad clínicamente importante. No obstante, este análisis incorrecto es conservador, ya que los estudios reciben una ponderación menor, en lugar de una ponderación exagerada. Aunque algunos argumentan en contra de la inclusión de los ensayos cruzados de esta forma, el error de unidad de análisis se pudiera considerar menos grave que otros tipos de errores de unidad de análisis.

Un segundo enfoque para incorporar los ensayos cruzados es incluir solamente los datos del primer período. Esto pudiera ser apropiado si se considera que el efecto de arrastre es un problema, o si se considera que un diseño cruzado no es apropiado por otros motivos. Sin embargo, es posible que los datos disponibles de los primeros períodos sean un subgrupo sesgado de los datos de todos los primeros períodos, lo que se debería a que la publicación de los datos del primer período puede depender de que los autores del ensayo hayan encontrado un efecto de arrastre estadísticamente significativo.

Un tercer enfoque para incorporar ensayos cruzados informados de forma inapropiada es intentar aproximar un análisis pareado al imputar las desviaciones estándar faltantes. Este enfoque se aborda en detalle en la Sección 16.4.6.

Los ensayos cruzados con resultados dicotómicos requieren métodos más complicados y se recomienda consultar a un estadístico (Elbourne 2002).

16.4.6 Análisis aproximados de los ensayos cruzados para un metanálisis

Introducción

La Tabla 16.4.a presenta algunos resultados que pudieran estar disponibles en la publicación de un ensayo cruzado, así como la notación que se utilizará en las secciones posteriores. Aquí se examinan métodos sencillos para aproximar análisis apropiados de los ensayos cruzados con el fin de obtener diferencias de medias o diferencias de medias estandarizadas para su uso en el metanálisis. Los revisores deberían considerar si la imputación de los datos faltantes es preferible a la exclusión completa de los ensayos cruzados de un metanálisis. La decisión dependerá de la confianza que puedan tener en los números imputados y en la solidez del resultado del metanálisis en un rango de resultados plausibles imputados.

Tabla 16.4.a: Algunos datos posibles disponibles en la publicación de un ensayo cruzado

Datos relacionados con	Estadísticos fundamentales	Estadísticos relacionados notificados frecuentemente
Intervención E	N, M_E, DE_E	Error estándar de M_E.
Intervención C	N, M_C, DE_C	Error estándar de M_C.
Diferencia entre E y C	N, DM, DE_{dif}	Error estándar de la DM; Intervalo de confianza para la DM; Estadística t pareada; Valor de p de la prueba t pareada.

16.4.6.1 Diferencias de medias

Habitualmente la estimación puntual de la diferencia de medias está disponible para un análisis pareado, ya que es la misma que para un análisis de grupos paralelos (la media de las diferencias es igual a la diferencia en las medias):

$$DM = M_E - M_C.$$

El error estándar de la diferencia de medias se obtiene como

$$SE(MD) = \frac{SD_{diff}}{\sqrt{N}}.$$

donde N es el número de participantes en el ensayo, y DE_{diff} es la desviación estándar de las *diferencias dentro de los participantes entre las mediciones E y C*. Como se indica en la Sección 16.4.4, el error estándar también se puede obtener directamente de un intervalo de confianza para la DM, de un estadístico t pareado o de un valor de p de una prueba t pareada. Las cantidades DM y EE (DM) se pueden introducir en RevMan bajo el tipo de resultado varianza inversa genérica.

Cuando el error estándar no está disponible directamente y no se presenta la desviación estándar de las diferencias, un enfoque simple es imputar la desviación estándar, como se hace frecuentemente para otras desviaciones estándar faltantes (ver Sección 16.1.3). Otros estudios en el metanálisis pueden presentar las desviaciones estándar de las diferencias, y siempre que el estudio utilice la misma escala de medición, puede ser razonable obtenerlas de un estudio para otro. Al igual que con las imputaciones, se deberían realizar análisis de sensibilidad para evaluar la repercusión de los datos imputados sobre los hallazgos del metanálisis (ver Sección 16.1 y Capítulo 9, Sección 9.7)

Si no hay información disponible de algún estudio sobre las desviaciones estándar de las diferencias, es posible lograr la imputación de las desviaciones estándar al suponer un coeficiente de correlación particular. El coeficiente de correlación describe cuán similares son las mediciones en las intervenciones E y C para un participante, y es un número entre -1 y 1. Es de esperar que el mismo se encuentre entre 0 y 1 en el contexto de un ensayo cruzado, ya que un resultado mayor que el promedio para un participante mientras reciba E tenderá a asociarse con un resultado mayor que el promedio mientras reciba C. Si el coeficiente de correlación es cero o negativo, entonces no hay un beneficio estadístico con el uso de un diseño cruzado con respecto al uso de un diseño paralelo.

Una forma frecuente de presentar los resultados de un ensayo cruzado es como si el ensayo hubiera sido un ensayo de grupos paralelos, con las desviaciones estándar para cada intervención por separado (DE_E y DE_C; ver Tabla 16.4.a). La desviación estándar de las diferencias esperada se puede calcular mediante estas desviaciones estándar específicas y un coeficiente de correlación imputado (Corr):

$$DE_{dif} = \sqrt{DE_E^{\,2} + DE_C^{\,2} \quad (2 \text{ x Corr x } DE_E \text{ x } DE_C)}\ .$$

16.4.6.2 Diferencia de medias estandarizada

La diferencia de medias estandarizada (DME) más apropiada de un ensayo cruzado divide la diferencia de medias por la desviación estándar de las mediciones (y no por la desviaciones estándar de las diferencias). Una DME se puede calcular a partir de las desviaciones estándar agrupadas específicas para la intervención de la siguiente manera:

$$SMD = \frac{MD}{SD_{pooled}},$$

donde

$$SD_{pooled} = \sqrt{\frac{SD_E^2 + SD_C^2}{2}}.$$

Se necesita un coeficiente de correlación para el error estándar de la DME:

$$SE(SMD) = \sqrt{\frac{1}{N} + \frac{SMD^2}{2N}} \times \sqrt{2(1 - Corr)} \ .$$

Alternativamente, la DME se puede calcular de la DM y su error estándar mediante una correlación imputada:

$$SMD = \frac{MD}{SE(MD) \times \sqrt{\frac{N}{2(1 - Corr)}}}$$

En este caso la correlación imputada repercute sobre la magnitud de la propia estimación del efecto de la DME (en lugar de sobre el error estándar solamente, como es el caso para los análisis de la DM en la Sección 16.4.6.1). Por lo tanto, las correlaciones imputadas se deberían utilizar con gran precaución para la estimación de las DME.

16.4.6.3 Imputación de los coeficientes de correlación

El valor para un coeficiente de correlación se pudiera imputar de otro estudio en el metanálisis (ver más adelante), de una fuente externa al metanálisis o se pudiera hipotetizar a partir de un argumento razonado. En todas estas situaciones se debería realizar un análisis de sensibilidad en el que se prueben los diferentes valores de Corr para determinar si el resultado general del análisis es sólido con respecto al uso de los coeficientes de correlación imputados.

Es posible la estimación de un coeficiente de correlación a partir de otro estudio del metanálisis si dicho estudio presenta las tres desviaciones estándar de la Tabla 16.4.a. El cálculo supone que la media y la desviación estándar de las mediciones para la intervención E son las mismas cuando dicha intervención se administra en el primer período que cuando se administra en el segundo período (y de manera similar para la intervención C).

$$Corr = \frac{SD_E^{\ 2} + SD_C^{\ 2} - SD_{diff}^{\ 2}}{2 \times SD_E \times SD_C} \ .$$

Antes de realizar la imputación se recomienda computar los coeficientes de correlación para cuantos estudios sea posible y compararlos. Si estas correlaciones varían de manera significativa entonces los análisis de sensibilidad son particularmente importantes.

16.4.6.4 Ejemplo

Como ejemplo, suponga un ensayo cruzado que presente los siguientes datos:

Intervención E (tamaño de muestra 10)	M_E = 7,0 DE_E = 2,38
Intervención C (tamaño de muestra 10)	M_C = 6,5 DE_C = 2,21

Diferencia de medias, imputación de la DE de las diferencias (DE_{diff})

La estimación de la diferencia de medias es DM = 7,0 – 6,5 = 0,5. Suponga que la desviación estándar típica que se ha observado en otros ensayos es 2. Entonces es posible calcular el error estándar de la DM como

$$EE\ (DM) = \frac{DE_{dif}}{\sqrt{N}} = \frac{2}{\sqrt{10}} = 0{,}632.$$

Los números 0,5 y 0,632 se pueden introducir en RevMan como la estimación y el error estándar de una diferencia de medias, como una variable genérica del inverso de la varianza.

Diferencia de medias, imputación del coeficiente de correlación (Corr)

La estimación de la diferencia de medias es nuevamente DM = 0,5. Suponga que se ha imputado un coeficiente de correlación de 0,68. Entonces es posible imputar la desviación estándar de la diferencia como:

$$DE_{dif} = \sqrt{DE_E^{\ 2} + DE_C^{\ 2} - (2 \text{ x } Corr \text{ x } DE_E \text{ x } DE_C)}$$

$$= \sqrt{2{,}38^2 + 2{,}21^2 - (2 \text{ x } 0{,}68 \text{ x } 2{,}38 \text{ x } 2{,}21)} \quad 1{,}8426$$

Entonces el error estándar de la DM es

$$EE\ (DM) = \frac{DE_{dif}}{\sqrt{N}} = \frac{1{,}8426}{\sqrt{10}} = 0{,}583\ .$$

Los números 0,5 y 0,583 se pueden introducir en RevMan como la estimación y el error estándar de una diferencia de medias como una variable genérica del inverso de la varianza. Los coeficientes de correlación diferentes de 0,68 se deberían utilizar como parte de un análisis de sensibilidad.

Diferencia de medias estandarizada, imputación del coeficiente de correlación (Corr)

La diferencia de medias estandarizada se puede calcular directamente de los datos:

$$DME = \frac{DM}{DE_{agrupados}} = \frac{DM}{\sqrt{\frac{DE_E^{\ 2} + DE_C^{\ 2}}{2}}} = \frac{0{,}5}{\sqrt{\frac{2{,}38^2 + 2{,}21^2}{2}}} = 0{,}218.$$

Por lo tanto, el error estándar se obtiene como:

$$EE(DME) = \sqrt{\frac{1}{N} + \frac{DME^2}{2N}} \text{x} \sqrt{2(1 - Corr} = \sqrt{\frac{1}{10} + \frac{0{,}218^2}{20}} \text{x} \sqrt{2(1 - 0{,}68)} \quad 0{,}256\ .$$

Los números 0,218 y 0,256 se pueden introducir en RevMan como la estimación y el error estándar de una diferencia de medias estandarizada como una variable genérica del inverso de la varianza.

La DME también se puede obtener de la DM y su error estándar:

$$DME = \frac{DM}{EE(DM) \text{ x } \sqrt{\frac{N}{2(1 - Corr)}}} = \frac{0{,}5}{0{,}583 \text{ x } \sqrt{\frac{10}{2(1 - 0{,}68)}}} = 2{,}17$$

La pequeña discrepancia surge debido a las formas ligeramente diferentes en las cuales las dos fórmulas calculan una desviación estándar agrupada para la estandarización.

16.4.7 Problemas en la incorporación de los ensayos cruzados

En principio, los ensayos cruzados se pueden combinar con ensayos de grupos paralelos en el mismo metanálisis. Se debería tener en cuenta la posibilidad de diferencias importantes en otras características entre los diferentes tipos de ensayos. Por ejemplo, los ensayos cruzados pueden tener períodos más cortos de intervención o pueden incluir participantes con enfermedad menos grave. En general es recomendable realizar un metanálisis de los ensayos de grupos paralelos y cruzados por separado, independientemente de si también se combinan juntos.

Los revisores deberían señalar explícitamente cómo han procesado los datos de los ensayos cruzados y deberían realizar análisis de sensibilidad para investigar la solidez de sus conclusiones, especialmente cuando los coeficientes de correlación se han obtenido de fuentes externas (ver Capítulo 9, Sección 9.7). Se recomienda buscar apoyo estadístico.

16.5 Estudios con más de dos grupos de intervención

16.5.1 Introducción

No es poco frecuente que los ensayos clínicos asignen al azar a los participantes a uno de varios grupos de intervención. Una revisión de ensayos aleatorizados publicada en diciembre de 2000 encontró que la cuarta parte tuvo más de dos grupos de intervención (Chan 2005). Por ejemplo, puede haber dos o más grupos de intervención experimental con un grupo control común, o dos grupos de intervención control como un grupo placebo y un grupo de tratamiento estándar. A estos estudios se les llama estudios “de brazos múltiples”. Un caso especial es un ensayo factorial, el cual aborda dos o más comparaciones de intervenciones simultáneas mediante cuatro o más grupos de intervención (ver Sección 16.5.6).

Aunque una revisión sistemática puede incluir varias comparaciones de intervenciones (y por lo tanto varios metanálisis), casi todos los metanálisis abordan comparaciones pareadas. Existen tres aspectos separados a considerar cuando se encuentra un estudio con más de dos grupos de intervención:

1. Determinar qué grupos de intervención son relevantes para la revisión sistemática.
2. Determinar qué grupos de intervención son relevantes para un determinado metanálisis.
3. Determinar cómo se incluirá el estudio en el metanálisis si más de dos grupos son relevantes.

16.5.2 Determinación de qué grupos de intervención son relevantes

Para un determinado estudio con brazos múltiples los grupos de intervención de relevancia para una *revisión sistemática* son todos los que se pudieran incluir en una comparación pareada de grupos de intervención que, si se investigan solos, cumplirían los criterios de inclusión de los estudios en la revisión. Por ejemplo, una revisión que aborde solamente una comparación de “terapia de reemplazo con nicotina versus placebo” para el abandono del

hábito de fumar pudiera identificar un estudio que compare “goma de mascar de nicotina versus terapia conductual versus goma de mascar placebo”. De las tres posibles comparaciones de intervenciones pareadas sólo una (“goma de mascar de nicotina versus goma de mascar placebo) aborda el objetivo de la revisión y ninguna de las comparaciones que incluyen la terapia conductual lo hace. Por lo tanto, el grupo de terapia conductual no es relevante para la revisión. Sin embargo, si el estudio ha comparado “goma de mascar de nicotina más terapia conductual versus terapia conductual más goma de mascar placebo versus goma de mascar placebo sola”, una comparación de las dos primeras intervenciones se pudiera considerar relevante, no así el grupo de goma de mascar placebo.

Como ejemplo de grupos control múltiples, una revisión que aborda la comparación “acupuntura versus ninguna acupuntura” pudiera identificar un estudio que compare “acupuntura versus acupuntura simulada versus ninguna intervención”. Los revisores se preguntarían si, por una parte, un estudio que compare “acupuntura versus acupuntura simulada” se incluiría en la revisión y, por otra parte, si un estudio de “acupuntura versus ninguna intervención” se incluiría. Si ambos se incluyen, entonces los tres grupos de intervención del estudio son relevantes para la revisión.

Como regla general y para evitar confusiones al lector acerca de la identidad y la naturaleza de cada estudio, se recomienda que todos los grupos de intervención de un estudio de intervenciones múltiples se mencionen en la tabla “Características de los estudios incluidos”, en la celda “Intervenciones” o en la celda “Notas”. Sin embargo, es necesario proporcionar descripciones detalladas solamente de los grupos de intervención relevantes para la revisión, y sólo se deberían utilizar esos grupos en los análisis.

Las mismas consideraciones de relevancia se aplican cuando se determina qué grupos de intervención de un estudio se deberían incluir en un determinado *metanálisis*. Cada metanálisis aborda solamente una comparación pareada única, por lo que los revisores deberían considerar si un estudio de cada comparación pareada posible de las intervenciones del estudio sería elegible para el metanálisis. Para distinguir entre la decisión a nivel de revisión y la decisión a nivel de metanálisis considere una revisión de “terapia con nicotina versus placebo u otros comparadores”. Todos los grupos de intervención de un estudio de “goma de mascar de nicotina versus terapia conductual versus goma de mascar placebo” pudieran ser relevantes para la revisión. Sin embargo, es posible que la presencia de múltiples intervenciones no represente un problema para los metanálisis, ya que es probable que “goma de mascar de nicotina versus goma de mascar placebo” y “goma de mascar de nicotina versus terapia conductual” se aborden en metanálisis diferentes. Por el contrario, todos los grupos del estudio “acupuntura versus acupuntura simulada versus ninguna intervención” se pudieran considerar elegibles para el mismo metanálisis si éste incluye un estudio de “acupuntura versus acupuntura simulada” y un estudio de “acupuntura versus ninguna intervención”. En la Sección 16.5.4 se describen los métodos para tratar esta última situación.

16.5.3 Evaluación del riesgo de sesgo en los estudios con más de dos grupos

Es posible introducir sesgo en un estudio de intervenciones múltiples si las decisiones con respecto al análisis de los datos se toman después de ver los datos. Por ejemplo, los grupos que reciben diferentes dosis de la misma intervención se pueden combinar sólo después de ver los resultados, incluidos los valores de p. Además, es posible presentar diferentes resultados cuando se comparan diferentes pares de grupos, que nuevamente es posible que estén relacionados con los hallazgos.

Juszczak y cols. examinaron 60 ensayos aleatorizados de intervenciones múltiples, de los cuales cerca de la tercera parte tenían al menos cuatro brazos de intervención (Juszczak 2003). Encontraron que sólo el 64% presentó las mismas comparaciones de grupos para todos los resultados, lo que indica una notificación selectiva similar a la notificación selectiva de un ensayo de dos brazos. Además, el 20% presentó la combinación de grupos en un análisis. Sin embargo, si se proporcionan los datos resumidos para cada grupo de intervención, sin importar cómo se han combinado los grupos en los análisis informados, los revisores no necesitan analizar los datos de la misma manera que lo hicieron los autores del estudio.

Algunas preguntas indicadas para evaluar el riesgo de sesgo en los estudios con intervenciones múltiples son las siguientes:

- ¿Los datos se presentan para cada uno de los grupos a los cuales los participantes se asignaron al azar?
- ¿Las descripciones del estudio están libres de sospecha de notificación selectiva de las comparaciones de los brazos de intervención para algunos resultados?

Si la respuesta a la primera pregunta es “sí” la segunda pregunta no es importante (por lo que la misma también se pudiera responder como “sí”).

16.5.4 Cómo incluir múltiples grupos de un estudio

Hay muchos enfoques posibles para incluir un estudio con múltiples grupos de intervención en un determinado metanálisis. Un enfoque que se debería evitar es sencillamente introducir numerosas comparaciones en el metanálisis cuando las mismas tienen uno o más grupos de intervención en común. Este procedimientos provoca un “doble recuento” de los participantes en el/los grupo/s de intervención compartido/s y crea un error de unidad de análisis debido a una correlación no abordada entre las estimaciones de los efectos de la intervención de comparaciones múltiples (ver Capítulo 9, Sección 9.3). Es importante diferenciar entre las situaciones en las cuales un estudio puede contribuir con varias comparaciones *independientes* (es decir, sin un grupo de intervención en común) y cuando varias comparaciones se *correlacionan* debido a que tienen grupos de intervención y, por lo tanto, participantes en común. Por ejemplo, considere un estudio que asignó al azar a los participantes a cuatro grupos: “goma de mascar de nicotina” versus “goma de mascar placebo” versus “parche de nicotina” versus “parche placebo”. Un metanálisis que aborde la pregunta amplia de si la terapia de reemplazo con nicotina es efectiva pudiera incluir la comparación “goma de mascar de nicotina versus goma de mascar placebo” así como la comparación independiente “parche de nicotina versus parche placebo”. Habitualmente es razonable incluir comparaciones independientes en un metanálisis como si provinieran de estudios diferentes, aunque hay complicaciones sutiles con respecto a los análisis de efectos aleatorios (ver Sección 16.5.5).

Los enfoques para sortear un error de unidad de análisis para un estudio que pudiera contribuir con comparaciones múltiples correlacionadas incluyen los siguientes:

- Combinar grupos para crear una comparación única pareada (recomendado).
- Seleccionar un par de intervenciones y excluir las otras.
- Separar el grupo “compartido” en dos o más grupos con tamaños de muestra más pequeños, e incluir dos más comparaciones (razonablemente independientes).
- Incluir dos o más comparaciones correlacionadas y tener en cuenta la correlación.
- Realizar un metanálisis de tratamientos múltiples (ver Sección 16.6).

El método recomendado en la mayoría de las situaciones es combinar todos los grupos de intervención experimental relevantes del estudio en un grupo único y combinar todos los grupos de intervención control relevantes en un grupo control único. Como ejemplo, suponga que un metanálisis de “acupuntura versus no acupuntura” consideraría que los estudios de “acupuntura versus acupuntura simulada” o los estudios de “acupuntura versus ninguna intervención” son elegibles para inclusión. Entonces un estudio que compare “acupuntura versus acupuntura simulada versus ninguna intervención” se incluiría en el metanálisis al combinar los participantes del grupo “acupuntura simulada” con los participantes del grupo “ninguna intervención”. Este grupo control combinado se compararía con el grupo “acupuntura” de la forma habitual. Para los resultados dicotómicos es posible sumar los tamaños de muestra y el número de personas con eventos entre los grupos. Para los resultados continuos es posible combinar las medias y las desviaciones estándar mediante los métodos descritos en el Capítulo 7 (Sección 7.7.3.8).

La estrategia alternativa de seleccionar un par único de intervenciones (p.ej. seleccionar “acupuntura simulada” o “ninguna intervención” como el control) da lugar a pérdida de información y está abierta a selecciones relacionadas con los resultados, por lo que en general no se recomienda.

Una posibilidad adicional es incluir cada comparación pareada por separado, pero con los grupos de intervención compartidos divididos de forma aproximadamente igual entre las comparaciones. Por ejemplo, si un ensayo compara 121 pacientes que reciben acupuntura con 124 pacientes que reciben acupuntura simulada y 117 pacientes que no reciben acupuntura, entonces se pudieran introducir dos comparaciones en el metanálisis (de, por ejemplo, 61 “acupuntura” contra 124 “acupuntura simulada”, y de 60 “acupuntura” contra 117 “ninguna intervención”). Para los resultados dicotómicos se dividirían el número de eventos y el número total de pacientes. Para los resultados continuos sólo se dividiría el número total de participantes y las medias y las desviaciones estándar se mantendrían sin cambio. Este método resuelve sólo parcialmente el error de unidad de análisis (porque las comparaciones resultantes permanecen correlacionadas), por lo que en general no se recomienda. Sin embargo, una ventaja potencial de este enfoque sería que es posible realizar investigaciones aproximadas de la heterogeneidad entre los brazos de intervención (por ejemplo, en el caso del ejemplo presentado aquí, la diferencia entre el uso de la acupuntura simulada y ninguna intervención como grupo control).

Dos opciones finales, que necesitarían apoyo estadístico, son tener en cuenta la correlación entre las comparaciones correlacionadas del mismo estudio en el análisis, y realizar un metanálisis de tratamientos múltiples. La primera incluye calcular un promedio (o un promedio ponderado) de las comparaciones pareadas relevantes del estudio, calcular una varianza (y por lo tanto una ponderación) para el estudio y tener en cuenta la correlación entre las comparaciones. Este procedimiento producirá típicamente un resultado similar al método recomendado de combinar los grupos experimental y control. El metanálisis de tratamientos múltiples se discute con más detalle en la Sección 16.6.

16.5.5 Consideraciones sobre la heterogeneidad con estudios de intervenciones múltiples

Dos posibilidades para abordar la heterogeneidad entre los estudios es tenerla en cuenta en un metanálisis de efectos aleatorios, e investigarla mediante análisis de subgrupos o metarregresión (Capítulo 9, Sección 9.6). Algunas complicaciones surgen cuando se incluyen estudios de intervenciones múltiples en dichos análisis. Primero, no será posible investigar ciertas fuentes de heterogeneidad relacionadas con la intervención si los grupos de

intervención se combinan como el enfoque recomendado en la Sección 16.5.4. Por ejemplo no es posible subagrupar según “acupuntura simulada” o “ninguna intervención” como control si estos dos grupos se combinan antes del metanálisis. La forma más sencilla de realizar una investigación de esta diferencia entre los estudios es crear dos o más comparaciones del estudio (p.ej. “acupuntura versus acupuntura simulada” y “acupuntura versus ninguna intervención”). Sin embargo, si dichas comparaciones contienen un grupo de intervención común (como aquí, el grupo acupuntura) entonces no son independientes y ocurrirá un error de unidad de análisis, incluso si el tamaño de muestra se reduce debido al/los grupo/s de intervención compartido/s. No obstante, la división del tamaño de muestra para el grupo de intervención compartido aún es una forma práctica de realizar investigaciones aproximadas de la heterogeneidad.

Un problema más sutil ocurre en los metanálisis de efectos aleatorios si se incluyen comparaciones múltiples del mismo estudio. Un metanálisis de efectos aleatorios tiene en cuenta la variación al suponer que los efectos de los estudios en el metanálisis siguen una distribución entre ellos. La intención es tener en cuenta la variación de estudio a estudio. Sin embargo, si dos o más estimaciones provienen del mismo estudio, se espera la misma variación entre las comparaciones dentro del estudio y entre los estudios. Este hecho es cierto si las comparaciones son independientes o están correlacionadas (ver Sección 16.5.4). Una forma de sortear esta dificultad es realizar un metanálisis de efectos fijos entre las comparaciones dentro del estudio, y un metanálisis de efectos aleatorios entre los estudios. Se recomienda apoyo estadístico; en la práctica es probable que la diferencia entre varios análisis sea trivial.

16.5.6 Ensayos factoriales

En un ensayo factorial se realizan dos (o más) comparaciones de intervenciones simultáneamente. Entonces, por ejemplo, los participantes se pueden asignar al azar a recibir aspirina o placebo, y también se asignan al azar a recibir una intervención conductual o una atención estándar. La mayoría de los ensayos factoriales tienen dos “factores” de esta manera, cada uno de ellos tiene dos niveles; a éstos se les llama ensayos factoriales 2x2. Ocasionalmente es posible encontrar ensayos factoriales 3x2, o ensayos que investigan tres, cuatro o más intervenciones simultáneamente. A menudo sólo una de las comparaciones será relevante para una determinada revisión. Los siguientes planteamientos se centran en el caso 2x2, pero los principios se extienden a diseños más complejos.

En la mayoría de los ensayos factoriales la intención es lograr “dos ensayos por el precio de uno”, y se hace la suposición de que los efectos de las diferentes intervenciones activas son independientes, es decir, no hay interacción (sinergia). Ocasionalmente es posible realizar un ensayo específicamente para investigar si hay interacción entre dos tratamientos. Este aspecto se puede explorar con mayor frecuencia en un ensayo que compare cada uno de los dos tratamientos activos por separado con ambos combinados, sin un grupo placebo. Dichos ensayos no son factoriales.

El diseño factorial 2x2 se puede mostrar como una tabla 2x2, donde las filas indican una comparación (p.ej. aspirina versus placebo) y las columnas indican la otra (p.ej. intervención conductual versus atención estándar):

		Asignación al azar a B	
		Intervención conductual (B)	Atención estándar (no B)
Asignación al azar a A	Aspirina (A)	A y B	A, no B
	Placebo (no A)	B, no A	No A, no B

Un ensayo factorial se puede ver como dos ensayos que abordan diferentes preguntas. Es importante que ambas partes del ensayo se describan como si fueran solamente un ensayo de grupos paralelos de dos brazos. Por lo tanto, es de esperar ver los resultados para aspirina versus placebo, en los que se incluyan todos los participantes independientemente de si recibieron una intervención conductual o atención estándar, y de la misma manera para la intervención conductual. Estos resultados se pueden observar relacionados con los márgenes de la tabla 2x2. También sería conveniente evaluar si existe cierta interacción entre los tratamientos (es decir, el efecto de A depende de si se recibió B o "no B"), para lo cual es necesario ver las cuatro celdas en la tabla (McAlister 2003). Esto se debería a que la práctica de publicar dos artículos separados, posiblemente en diferentes revistas, no permite observar los resultados completos.

McAlister y cols. revisaron 44 ensayos factoriales publicados (McAlister 2003). Encontraron que solamente el 34% proporcionó resultados para cada celda de la estructura factorial. Sin embargo, habitualmente será posible derivar los resultados marginales de los resultados de las cuatro celdas en la estructura 2x2. En la misma revisión el 59% de las publicaciones de los ensayos incluyeron resultados de una prueba de interacción. En el reanálisis, 2/44 ensayos (6%) tuvieron un valor de $p < 0,05$, que es cercano a lo esperado por el azar (McAlister 2003). Por lo tanto, a pesar de las inquietudes acerca de las interacciones no reconocidas, al parecer los investigadores limitan apropiadamente el uso del diseño factorial a las situaciones en las que dos (o más) tratamientos no tienen el potencial para una interacción significativa. Desafortunadamente muchos revisores no aprovechan este hecho y sólo incluyen la mitad de los datos disponibles en su metanálisis (p.ej. incluyen solamente los resultados de A versus no A entre los que *no* recibieron B, y excluyen la investigación válida de A entre los que *recibieron* B).

Una pregunta indicada para evaluar el riesgo de sesgo en los ensayos factoriales es la siguiente:

- ¿Las publicaciones del estudio están libres de indicio de una interacción importante entre los efectos de las diferentes intervenciones?

16.6 Comparaciones indirectas y metanálisis de tratamientos múltiples

16.6.1 Introducción

Las comparaciones directas de intervenciones alternativas pueden ser el foco de una revisión Cochrane de intervención, un objetivo secundario de una revisión Cochrane de intervención o un aspecto clave de una revisión de revisiones Cochrane. Las revisiones de revisiones

Cochrane sintetizan múltiples revisiones Cochrane, típicamente de diferentes intervenciones para la misma afección (ver Capítulo 22). Idealmente las comparaciones directas de intervenciones alternativas se harían dentro de los estudios aleatorizados, pero a menudo dichos estudios no están disponibles. Las comparaciones indirectas son comparaciones que se hacen entre intervenciones opuestas que no se han comparado directamente entre sí (ver Sección 16.6.2). El metanálisis de tratamientos múltiples (MTM) es una ampliación de las comparaciones indirectas que permite la combinación de comparaciones directas e indirectas, así como el análisis simultáneo de los efectos comparativos de muchas intervenciones (ver Sección 16.6.3).

16.6.2 Comparaciones indirectas

Las comparaciones indirectas se hacen entre intervenciones a falta de estudios aleatorizados directos. Por ejemplo, suponga que algunos ensayos han comparado la efectividad de "dietista versus médico/a" para proporcionar asesoramiento dietético, y otros han comparado la efectividad de "dietista versus enfermero/a", pero ninguno ha comparado la efectividad de "médico/a versus enfermero/a". Pudiera ser conveniente entonces conocer la efectividad relativa de "médico/a versus enfermero/a" mediante comparaciones indirectas. De hecho, los profesionales de la medicina y de la enfermería se pueden comparar de forma indirecta al contrastar los ensayos de "dietista versus médico/a" con los de "dietista versus enfermero/a".

Un enfoque que nunca se debería utilizar es la comparación directa de los brazos únicos relevantes de los ensayos. Por ejemplo, los pacientes que reciben asesoramiento de un profesional de la enfermería (en los ensayos "dietista versus enfermero/a") no se deberían comparar directamente con los pacientes que reciben asesoramiento de un profesional de la medicina (en los ensayos "dietista versus médico/a"). Esta comparación ignora los beneficios potenciales de la asignación al azar y presenta los mismos sesgos (habitualmente grandes) de una comparación de estudios de cohortes independientes.

Hay métodos de comparación indirecta más apropiados disponibles, pero se deberían considerar cuidadosamente las suposiciones que subyacen bajo estos métodos. Un método relativamente sencillo es realizar análisis de subgrupos, donde los diferentes subgrupos se definen por las diferentes comparaciones realizadas. Para el caso particular de dos subgrupos (dos comparaciones, tres intervenciones) es posible calcular las diferencias entre los subgrupos y determinar la significación estadística mediante un procedimiento sencillo descrito por Bucher (Bucher 1997). En el ejemplo anterior, un subgrupo serían los ensayos "dietista versus médico/a", y el otro subgrupo serían los ensayos "dietista versus enfermero/a". La diferencia entre los efectos resumidos en los dos subgrupos proporcionará una estimación de la comparación deseada, "médico/a versus enfermero/a". Es posible utilizar la prueba para las diferencias entre subgrupos implementada en RevMan (ver Capítulo 9, Sección 9.6.3.1). La validez de una comparación indirecta se basa en que los diferentes subgrupos de los ensayos son similares, como promedio, en cuanto a todos los otros factores que pueden afectar el resultado. Existen discusiones más amplias de las comparaciones indirectas (Song 2003, Glenny 2005).

Las comparaciones indirectas no son comparaciones aleatorias, y no se pueden interpretar como tales. Son fundamentalmente hallazgos observacionales entre los ensayos, y pueden presentar los sesgos de los estudios observacionales, por ejemplo debido a factores de confusión (ver Capítulo 9, Sección 9.6.6). En situaciones en las que hay comparaciones directas e indirectas disponibles en una revisión, a menos que haya deficiencias en el diseño de los ensayos directos, los dos enfoques se deberían considerar por separado y las comparaciones directas deberían predominar como base para establecer conclusiones.

16.6.3 Metanálisis de tratamientos múltiples

Existen métodos disponibles para analizar simultáneamente tres o más intervenciones diferentes en un metanálisis. Habitualmente se les llama "metanálisis de tratamientos múltiples" ("MTM"), "metanálisis en red" o metanálisis de "comparaciones de tratamientos mixtos" ("CTM"). Los metanálisis de tratamientos múltiples se pueden utilizar para analizar estudios con múltiples grupos de intervención, y resumir estudios que hacen diferentes comparaciones de intervenciones. Caldwell y cols. proporcionan un introducción interesante (Cadwell 2005); Salanti y cols. (Salanti 2008) brindan una discusión más exhaustiva. Es de señalar que los metanálisis de tratamientos múltiples mantienen la identidad de cada intervención y permiten realizar comparaciones de múltiples intervenciones. Esto contrasta con los métodos para tratar un estudio único con múltiples grupos de intervención que se describen en la Sección 16.5, que se centran en reducir los múltiples grupos a una comparación única pareada.

El ejemplo más sencillo de un metanálisis de tratamientos múltiples es la comparación indirecta descrita en la Sección 16.6.2. Con tres intervenciones (p.ej. asesoramiento de un dietista, de un profesional de la medicina y uno de la enfermería), se pueden comparar indirectamente cualquiera de dos con una tercera. Por ejemplo, los profesionales de la medicina y la enfermería se pueden comparar indirectamente al contrastar los ensayos de "dietista versus médico/a" con los ensayos de "dietista versus enfermero/a". Este análisis se puede ampliar de varias formas. Por ejemplo, si también existen ensayos de comparaciones directas "médico/a versus enfermero/a" éstos se pudieran combinar con los resultados de la comparación indirecta. Si existen más de tres intervenciones habrá varias comparaciones directas e indirectas, y será más conveniente analizarlas simultáneamente.

Si cada estudio compara exactamente dos intervenciones es posible realizar un metanálisis de tratamientos múltiples mediante análisis de subgrupos, y utilizar la prueba para diferencias de subgrupos como se describe en el Capítulo 9 (Sección 9.6.3.1). Sin embargo es preferible utilizar un modelo de efectos aleatorios para tener en cuenta la heterogeneidad dentro de cada subgrupo, lo que se puede lograr al utilizar en su lugar la metarregresión (ver Capítulo 9, Sección 9.6.4). Cuando algunos estudios incluyen más de dos grupos de intervención la síntesis requiere métodos de metanálisis multivariados. Ya no es posible utilizar el análisis estándar de subgrupos ni los métodos de metarregresión, aunque se puede realizar el análisis en un marco bayesiano mediante WinBUGS (ver Sección 16.8.1). Una ventaja particular de utilizar los marcos bayesianos es que todas las intervenciones en el análisis se pueden clasificar mediante métodos probabilísticos, en lugar de métodos brutos.

Los metanálisis de tratamientos múltiples son particularmente adecuados para los problemas abordados en Revisiones de revisiones (Capítulo 22). Sin embargo, estos se basan en una fuerte suposición de que los estudios de diferentes comparaciones son similares en todo, excepto en la intervención que se compara. Las comparaciones indirectas incluidas no son comparaciones aleatorias y pueden presentar los sesgos de los estudios observacionales, por ejemplo debido a los factores de confusión (ver Capítulo 9, Sección 9.6.6). En situaciones en las que hay disponibles comparaciones directas e indirectas en una revisión, el uso de cualquier metanálisis de tratamientos múltiples debería complementar, en lugar de sustituir, las comparaciones directas. Para un metanálisis de tratamientos múltiples se requiere apoyo estadístico experto, así como experiencia en el tema.

16.7 Multiplicidad y función del azar

16.7.1 Introducción

Una revisión Cochrane pudiera incluir análisis múltiples debido a la selección de varias medidas de resultado, a resultados medidos en diferentes puntos temporales, a la intención de realizar análisis de subgrupos, a la inclusión de comparaciones de intervenciones múltiples o a otros motivos. Mientras más análisis se realizan es más probable que se encuentre que alguno de ellos sea “estadísticamente significativo” sólo debido al azar. Al utilizar el nivel de significación convencional del 5%, es de esperar que una de cada 20 pruebas sea estadísticamente significativa, incluso cuando no haya una diferencia verdadera entre las intervenciones comparadas. Sin embargo, después de 14 pruebas independientes, es más probable que al menos una prueba sea significativa, en comparación con que no lo sea (probabilidad mayor de 0,05), incluso cuando no hay un efecto verdadero. La probabilidad de encontrar al menos un resultado estadísticamente significativo aumenta con el número de pruebas realizadas. La probabilidad de un hallazgo espurio por azar es mayor cuando los análisis son independientes. Por ejemplo, los análisis múltiples de diferentes subgrupos habitualmente son más problemáticos en este aspecto que los análisis múltiples de varios resultados, ya que estos últimos incluyen los mismos participantes, por lo que no son independientes.

El problema de la significación de la pruebas múltiples ocurre en los ensayos clínicos, y en la investigación en epidemiología y en salud pública (Bauer 1991, Ottenbacher 1998), así como en las revisiones sistemáticas (Bender 2008). Existe una amplia bibliografía estadística acerca del aspecto de la multiplicidad. Se han desarrollado varios enfoques estadísticos para realizar ajustes para las pruebas múltiples en varias situaciones (Bender 2001, Cook 2005, Dmitrienko 2006). Sin embargo, no hay consenso acerca de cuándo se debería tener en cuenta la multiplicidad ni acerca de qué enfoque estadístico se debería utilizar si se hace un ajuste para las pruebas múltiples. Por ejemplo, el uso de ajustes apropiados para pruebas independientes dará lugar a valores de p demasiado grandes cuando las pruebas múltiples no son independientes. Los ajustes para las pruebas múltiples se utilizan en ensayos clínicos confirmatorios para proteger contra las conclusiones significativas espurias cuando se utilizan múltiples pruebas de hipótesis (Koch 1996) y se han incorporado en las guías estadísticas correspondientes (CPMP Working Party on Efficacy of Medicinal Products 1995). En los estudios exploratorios, en los cuales no hay hipótesis clave preespecificadas, es posible que no sea necesario realizar ajustes para las pruebas múltiples y a menudo no son posibles (Bender 2001). Los resultados estadísticamente significativos de los estudios exploratorios se deberían considerar como “generadores de hipótesis”, independientemente de si se realizaron ajustes para las pruebas múltiples.

16.7.2 Multiplicidad en las revisiones sistemáticas

Los ajustes para pruebas múltiples no se utilizan habitualmente en las revisiones sistemáticas y en general no se recomienda su uso. No obstante, los aspectos de la multiplicidad se aplican por igual a las revisiones sistemáticas y a otros tipos de investigaciones. Los revisores deberían recordar que en una revisión Cochrane generalmente se debería hacer énfasis en la estimación de los efectos de la intervención en lugar de las pruebas para los mismos. Sin embargo, el problema general de las comparaciones múltiples afecta la estimación del intervalo tanto como a la prueba de hipótesis (Chen 2005, Bender 2008).

En las revisiones sistemáticas ocurren algunos problemas adicionales asociados con la multiplicidad. Por ejemplo, cuando se presentan los resultados de un estudio, no siempre es

posible conocer cuántas pruebas o análisis se realizaron. Es probable que en algunos estudios se seleccionen hallazgos interesantes para la presentación o la publicación relacionados con la significación estadística, y se omitan otros hallazgos "no interesantes", lo que da lugar a resultados erróneos y conclusiones espurias. La notificación selectiva de los resultados se discute con más detalle en el Capítulo 8 (Sección 8.13).

La planificación adecuada de las pruebas estadísticas de hipótesis (incluido cualquier ajuste para pruebas múltiples) se debería hacer idealmente en el estadio de diseño. Desafortunadamente, esto puede ser difícil para las revisiones sistemáticas cuando no se conoce al inicio qué resultados y qué medidas de efecto estarán disponibles de los estudios incluidos. Esta situación puede hacer más difícil e incluso imposible planificar *a priori* los procedimientos de múltiples pruebas para las revisiones sistemáticas. Además, sólo algunos procedimientos de las comparaciones múltiples desarrollados para estudios únicos se pueden utilizar en los metanálisis de los datos resumen. Se necesitan más investigación para desarrollar procedimientos adecuados para comparaciones múltiples para su uso en revisiones sistemáticas (Bender 2008).

En resumen, no hay una solución sencilla ni completamente satisfactoria para el problema de las pruebas múltiples y la estimación de intervalos múltiples en las revisiones sistemáticas. Sin embargo, es posible ofrecer el siguiente consejo general. Se pueden encontrar consejos más detallados en otras publicaciones (Bender 2008):

- En el protocolo de la revisión señalar qué análisis y qué resultados son de interés particular (mientras menos mejor). Los resultados se deberían clasificar previamente como resultados primarios y secundarios, y los resultados principales que aparezcan en la tabla "Resumen de los resultados" se deberían preespecificar. Si existe una hipótesis clave clara, que se debería probar mediante múltiples pruebas de significación, realizar un ajuste adecuado para pruebas múltiples brindará una confianza mayor en cualquier conclusión que se establezca.
- Aunque se recomienda que las revisiones Cochrane deberían tratar de incluir todos los resultados que es probable que sean importantes para los usuarios de la revisión, es más difícil establecer conclusiones generales si hay análisis múltiples. Hay que tener en cuenta, al establecer conclusiones, que aproximadamente una de cada 20 pruebas estadísticas independientes será estadísticamente significativa (a un nivel de significación del 5%), debido solamente al azar, cuando no hay una diferencia real entre los grupos.
- No seleccionar los resultados a enfatizar (p.ej. en el resumen) sobre la base de un valor de p estadísticamente significativo.
- Si hay varios puntos temporales a seleccionar para un resultado, se debería intentar presentar un efecto resumen del efecto general de todos los puntos temporales, o seleccionar el punto temporal que sea más apropiado (aunque la disponibilidad de datos adecuados de todos los ensayos puede ser un problema). Se deberían evitar las pruebas múltiples del efecto en cada uno de los puntos temporales.
- Mantenga los análisis de subgrupos al mínimo e interprételos con precaución.
- Interprete con precaución cualquier hallazgo que se haya hipotetizado previamente, incluso cuando sea "estadísticamente significativo". Estos hallazgos sólo se deberían utilizar para generar hipótesis, no para probarlas.

16.8 Enfoques bayesianos y jerárquicos para el metanálisis

16.8.1 Métodos bayesianos

Las estadísticas bayesianas son un enfoque estadístico basado en una filosofía diferente de la que fundamenta las pruebas de significación y los intervalos de confianza. Trata esencialmente de la actualización de la evidencia. En un análisis bayesiano la incertidumbre inicial se expresa a través de una distribución previa acerca de las cantidades de interés. Los datos y las suposiciones actuales con respecto a cómo se generaron se resumen en la probabilidad. La distribución posterior para las cantidades de interés se puede obtener al combinar la distribución previa y la probabilidad. La distribución posterior se puede resumir mediante estimaciones puntuales e intervalos creíbles, que se parecen mucho a las estimaciones y los intervalos de confianza clásicos. El análisis bayesiano no se puede realizar en RevMan, pero se puede realizar mediante el programa estadístico WinBUGS (smith 1995, Lunn 2000).

En el contexto de un metanálisis la distribución previa describirá la incertidumbre acerca de la medida de efecto particular que se analiza, como el *odds ratio* o la diferencia de medias. Esto puede ser una expresión de la creencia subjetiva acerca del tamaño del efecto, o puede provenir de diferentes fuentes de evidencia no incluidas en el metanálisis, como la información de estudios no aleatorizados. La amplitud de la distribución previa refleja el grado de incertidumbre acerca de la cantidad. Cuando existe poca o ninguna información, es posible utilizar una distribución “no informativa” previa, en la cual todos los valores entre el posible rango son igualmente probables. La probabilidad resume los datos de los estudios incluidos en el metanálisis (por ejemplo, tablas 2x2 de ensayos aleatorizados) y del modelo de metanálisis (por ejemplo, al suponer efectos fijos o efectos aleatorios).

La selección de la distribución previa es una fuente de controversia en las estadísticas bayesianas. Aunque es posible representar las creencias acerca de los efectos como una distribución previa, parece raro combinar datos objetivos de ensayos con opiniones subjetivas. Por lo tanto, una práctica común en el metanálisis es utilizar distribuciones previas no informativas para reflejar una posición de ignorancia previa. Esto es particularmente cierto para la comparación principal. Sin embargo, las distribuciones previas también pueden reemplazar otras cantidades en un metanálisis, como el grado de variación entre estudios en un análisis de efectos aleatorios. Puede ser útil utilizar una valoración o pruebas externas sobre algunos de estos otros parámetros, particularmente cuando hay pocos estudios en el metanálisis. Es importante realizar análisis de sensibilidad para investigar cómo dependen los resultados de cualquier suposición hecha.

Una diferencia entre el análisis bayesiano y el metanálisis clásico es que la interpretación se hace directamente en términos de creencia: un intervalo creíble del 95% para un *odds ratio* es la región en la cual se cree que se encuentra el *odds ratio* con una probabilidad del 95%. Así es como muchos médicos en realidad interpretan un intervalo de confianza clásico, pero en el marco estrictamente clásico el 95% se refiere a la frecuencia a largo plazo con la cual los intervalos del 95% contienen el valor verdadero. El marco bayesiano también permite que un revisor calcule la probabilidad de que el *odds ratio* tenga un rango particular de valores, lo cual no se puede realizar en el marco clásico. Por ejemplo, es posible determinar la probabilidad de que el *odds ratio* sea menor de 1 (lo que pudiera indicar un efecto beneficioso de una intervención experimental), o que no sea mayor de 0,8 (lo que pudiera indicar un efecto clínicamente importante). Se debería señalar que estas probabilidades son específicas para la selección de la distribución previa. Diferentes metanalistas pueden analizar los mismos datos mediante diferentes distribuciones previas y obtener diferentes resultados.

Los métodos bayesianos ofrecen algunas ventajas potenciales sobre muchos métodos clásicos para los metanálisis. Por ejemplo, se pueden utilizar para:

- incorporar pruebas externas, por ejemplo, sobre el efecto de las intervenciones o el grado probable de variación entre los estudios;
- ampliar un metanálisis a los contextos de toma de decisiones, al incorporar el concepto de la utilidad de varios estados de resultados clínicos;
- permitir de forma natural la imprecisión en la estimación de la varianza entre estudios (ver Capítulo 9, Sección 9.5.4);
- investigar la relación entre el riesgo subyacente y el beneficio del tratamiento (ver Capítulo 9, Sección 9.6.7);
- realizar análisis complejos (p.ej. metanálisis de tratamientos múltiples), debido a la flexibilidad del programa informático WinBUGS; y
- examinar el grado en el cual los datos cambiarían las creencias de las personas (Higgins 2002).

Se recomienda firmemente que los revisores que deseen realizar análisis bayesianos posean experiencia estadística. Hay numerosos textos de calidad (Sutton 2000, Sutton 2001, Spiegelhalter 2004).

16.8.2 Modelos jerárquicos

Algunas técnicas sofisticadas para el metanálisis explotan un marco estadístico llamado modelos jerárquicos, o modelos multinivel (Thompson 2001). Esto se debe a que la información en un metanálisis habitualmente proviene de dos niveles: los estudios en el nivel más alto, y los participantes dentro de los estudios en el nivel más bajo. En ocasiones los niveles adicionales pueden ser relevantes, por ejemplo los centros en un estudio multicéntrico o los grupos en un ensayo aleatorizado por grupos. Un marco jerárquico es apropiado si el metanálisis se realiza con información estadística resumida (por ejemplo, los logaritmos de los *odds ratios* y sus varianzas) o datos de pacientes individuales (Turner 2000). Este marco es particularmente relevante cuando se utilizan efectos aleatorios para representar la variación inexplicada en las estimaciones del efecto entre los estudios (ver Capítulo 9, Sección 9.5.4).

Los modelos jerárquicos, en lugar de los métodos de metanálisis más sencillos, son útiles en varios contextos. Por ejemplo, se pueden utilizar para:

- tener en cuenta la imprecisión de las estimaciones de la varianza de los efectos del tratamiento dentro de los estudios;
- tener en cuenta la imprecisión de la estimación de la varianza entre los estudios, tau cuadrado (ver Capítulo 9, Sección 9.5.4);
- proporcionar métodos para modelar explícitamente los datos de resultados binarios (en lugar de estadísticas resumen);
- investigar la relación entre el riesgo subyacente y el beneficio del tratamiento (ver Capítulo 9, Sección 9.6.7); y
- ampliar los métodos para incorporar las características a nivel de estudio (ver Capítulo 9, Sección 9.6.7) o las características a nivel individual (ver Capítulo 18)

Los modelos jerárquicos son particularmente relevantes cuando están disponibles los datos de pacientes individuales (DPI) sobre los resultados y las covariables (Higgins 2001). Sin

embargo, incluso si se utilizan estos métodos, es necesario tener precaución para asegurarse de que no se confunden las relaciones dentro y entre los estudios.

La implementación de los modelos jerárquicos necesita programas estadísticos sofisticados si se utiliza un enfoque estadístico clásico (p.ej. SAS proc mixed, o MlwiN) o un enfoque bayesiano (p.ej. WinBUGS). Muchas de las investigaciones metodológicas actuales en metanálisis utilizan métodos de modelos jerárquicos, a menudo con una implementación bayesiana.

16.9 Eventos poco frecuentes (incluidas las frecuencias cero)

16.9.1 Metanálisis de eventos poco frecuentes

Para los resultados poco frecuentes el metanálisis puede ser la única manera de obtener pruebas fiables de los efectos de las intervenciones en atención sanitaria. Habitualmente los estudios individuales no tienen poder estadístico suficiente para detectar diferencias en resultados poco frecuentes, pero el metanálisis de muchos estudios puede tener un poder estadístico adecuado para investigar si las intervenciones repercuten en la incidencia de eventos poco frecuentes. Sin embargo, muchos métodos de metanálisis se basan en aproximaciones de muestras grandes y no son adecuados cuando los eventos son poco frecuentes. Por lo tanto, los revisores deberían ser cuidadosos cuando seleccionan un método de metanálisis.

No hay un riesgo único para el cual los eventos se puedan clasificar como “poco frecuentes”. En realidad, riesgos de 1 en 1000 constituyen eventos poco frecuentes, y muchos clasificarían riesgos de 1 en 100 de la misma manera. Sin embargo, el rendimiento de los métodos cuando los riesgos son tan altos como 1 en 10 también se puede afectar por los problemas discutidos en esta sección. Lo que es típico es que una alta proporción de los estudios en el metanálisis no observe eventos en uno o más de los brazos del estudio.

16.9.2 Estudios con recuentos de celdas cero

Pueden ocurrir problemas computacionales cuando no se observan eventos en uno o ambos grupos en un estudio individual. Los métodos metanalíticos de la varianza inversa (los métodos de efectos fijos de la varianza inversa y los métodos de efectos aleatorios de DerSimonian y Laird) incluyen calcular una estimación del efecto de la intervención y su error estándar para cada estudio. Para los estudios en los que no se observan eventos en uno o ambos brazos estos cálculos a menudo incluyen dividir por un recuento cero, lo que produce un error de cálculo. La mayoría de los programas informáticos metanalíticos (incluido RevMan) verifica automáticamente los recuentos cero problemáticos y agrega un valor fijo (típicamente 0,5) a todas las celdas de las tablas de resultados del estudio en las que ocurre el problema. Los métodos de Mantel-Haenszel sólo requieren correcciones de celdas cero si la misma celda es cero en todos los estudios incluidos, y por lo tanto requieren el uso de corrección con menor frecuencia. Sin embargo, en muchas aplicaciones informáticas se aplican las mismas reglas de corrección para los métodos de Mantel-Haenszel que para los métodos de la varianza inversa. Los métodos del *odds ratio* y del cociente de riesgo requieren correcciones de las celdas cero con más frecuencia que los métodos de diferencias, excepto para el método de *odds ratio* de Peto, el cual tiene problemas de cálculo en la situación extrema en la cual no ocurren eventos en todos los brazos de todos los estudios.

Aunque la corrección fija cumple el objetivo de evitar errores de cálculo, habitualmente tiene el efecto indeseado de sesgar las estimaciones del estudio hacia la falta de diferencias y sobrestimar las varianzas de las estimaciones del estudio (y posteriormente disminuir de forma inapropiada la ponderación de su contribución al metanálisis). Cuando los tamaños del estudio son diferentes (lo que ocurre con mayor frecuencia en los ensayos no aleatorizados que en los aleatorizados), se introducirá un sesgo direccional en el efecto del tratamiento. Sweeting y cols. han explorado correcciones alternativas no fijas de celdas cero, incluida una corrección proporcional al recíproco del tamaño del brazo de estudio contrastante, y encontraron que era preferible a la corrección fija 0,5 cuando los tamaños de los brazos no están equilibrados (Sweeting 2004).

16.9.3 Estudios sin eventos

La práctica estándar en el metanálisis de *odds ratios* y riesgos relativos es excluir los estudios del metanálisis cuando no hay eventos en ambos brazos. Este enfoque se debe a que dichos estudios no proporcionan una indicación de la dirección ni la magnitud del efecto relativo del tratamiento. Aunque puede estar claro que los eventos son muy poco frecuentes en las intervenciones experimental y control, no se proporciona información acerca de qué grupo es más probable que tenga un riesgo mayor, o si los riesgos tienen el mismo o diferente orden de magnitud (cuando el riesgo es muy bajo, son compatibles con medidas de proporción muy grandes o muy pequeñas). Aunque uno se pudiera sentir tentado a inferir que el riesgo sería más bajo en el grupo con mayor tamaño de muestra (ya que el límite superior del intervalo de confianza sería menor), esto no se justifica, ya que los investigadores del estudio determinaron la asignación del tamaño de la muestra y no es una medida de la incidencia del evento.

Superficialmente los métodos de diferencias de riesgos parecen tener ventajas sobre los métodos de *odds ratio* en que la DR está definida (como cero) cuando no ocurren eventos en los brazos. Por lo tanto, estos estudios se incluyen en el proceso de estimación. Bradburn y cols. realizaron estudios de simulación que mostraron que todos los métodos de diferencias de riesgos producen intervalos de confianza demasiado amplios cuando los eventos son poco frecuentes, y se asocian con un poder estadístico deficiente, lo que los puede hacer inapropiados para el metanálisis de eventos poco frecuentes (Bradburn 2007). Esto es particularmente relevante cuando se estudian resultados que se centran en la seguridad del tratamiento, ya que la habilidad para identificar correctamente (o intentar refutar) los eventos adversos graves es un aspecto clave en el desarrollo de los fármacos.

Es probable que los resultados para los cuales no ocurren eventos en los brazos no se mencionen en las publicaciones de muchos ensayos aleatorizados, lo que impide su inclusión en el metanálisis. No obstante, no está claro cuando se trabaja con resultados publicados, si la falta de mención de un evento adverso particular significa que *no hubo* eventos, o sencillamente que dichos eventos no se incluyeron como una variable principal de resultado medida. Aunque los resultados de los metanálisis de las diferencias de riesgo estarán afectados por la falta de información de los resultados sin eventos, los métodos basados en los *odds ratios* y los riesgos relativos excluyen de forma natural estos datos estén o no publicados, por lo que no se afectarán.

16.9.4 Intervalos de confianza cuando no se observan eventos

Es posible colocar dentro de límites de confianza los riesgos de eventos cuando no se observan eventos, lo cual puede ser útil cuando se intenta confirmar los posibles riesgos de

eventos adversos graves. Se ha propuesto una regla sencilla llamada “regla de los tres”, de manera que si no se observan eventos en un grupo, el límite superior del intervalo de confianza para el número de eventos es tres, y para el riesgo (en una muestra de tamaño N) es 3/N (Hanley 1983). La aplicación de esta regla no se ha propuesto ni evaluado directamente para las revisiones sistemáticas. Sin embargo, cuando se examina la incidencia de un evento poco frecuente que no se observa en los grupos de intervención en una serie de estudios (ensayos aleatorizados, comparaciones no aleatorias o series de casos) parece razonable aplicarla, si se toma N como la suma de los tamaños de muestra de los brazos que reciben la intervención. No obstante, la misma no proporcionará información acerca de la incidencia relativa del evento ente los dos grupos.

El valor 3 coincide con el límite superior de un intervalo de confianza del 95% de una cola de la distribución de Poisson (equivalente al intervalo de confianza del 90% de dos colas). Para que el riesgo se encuentre en un intervalo de confianza más estándar del 97,5% de una cola (equivalente a un intervalo de confianza del 95% de dos colas) se debería utilizar 3,7 en todos los cálculos en lugar de 3 (Newcombe 2000). Una recomendación alternativa que produce valores similares es la “regla de los cuatro”, que toma como límite superior del riesgo 4/(N+4). Se recomienda el uso de cualquiera de estas opciones en las revisiones Cochrane. Por ejemplo, si no hubo eventos en 10 observaciones, el límite superior del intervalo de confianza para el número de eventos es 3,7 y para el riesgo es 3,7 de 10 (es decir 0,37). Si no hay eventos en 100 observaciones, el límite superior del número de eventos es todavía 3,7, pero el riesgo es 3,7 de 1000 (es decir 0,037).

16.9.5 Validez de los métodos de metanálisis para los eventos poco frecuentes

Los estudios de simulación han mostrado que muchos métodos metanalíticos pueden dar lugar a resultados erróneos para los eventos poco frecuentes, lo cual no es sorprendente debido a que se basan en la teoría estadística asintótica. Se ha considerado que su rendimiento es subóptimo porque los resultados están sesgados, los intervalos de confianza son inapropiadamente amplios o el poder estadístico es demasiado bajo para detectar diferencias significativas.

Más adelante se analiza la selección del método estadístico para los metanálisis de los *odds ratios*. Las decisiones apropiadas parecen depender del riesgo del grupo control, el tamaño probable del efecto del tratamiento y la consideración del equilibrio en los números de participantes tratados y control en los estudios incluidos. Ninguna investigación ha evaluado directamente las medidas del riesgo relativo, pero es probable que su rendimiento sea muy similar al correspondiente a la medición del *odds ratio*. Cuando los eventos son poco frecuentes las estimaciones de los odds y los riesgos son idénticas y los resultados de ambos se pueden interpretar como cocientes de probabilidades.

Bradburn y cols. encontraron que muchos de los métodos metanalíticos utilizados con más frecuencia estaban sesgados cuando los eventos fueron poco frecuentes (Bradburn 2007). El sesgo fue mayor en los métodos de la varianza inversa y del *odds ratio* y las diferencias de riesgos de DerSimonian y Laird, y en el método del *odds ratio* de Mantel-Haenszel que utilizan la corrección 0,5 para las celdas cero. Como ya se ha señalado, los métodos metanalíticos de las diferencias de riesgos tendieron a mostrar una cobertura más conservadora de los intervalos de confianza y un poder estadístico bajo cuando los riesgos de los eventos fueron bajos.

Con tasas de eventos por debajo del 1% se encontró que el método del *odds ratios* de un paso de Peto está menos sesgado y tiene un mayor poder estadístico, y que proporcionó la

mejor cobertura del intervalo de confianza, siempre que no hubiera un desequilibrio significativo entre el tamaño de los grupos de tratamiento y control dentro de los estudios y que los efectos del tratamiento no fueran excepcionalmente grandes. Este hallazgo se observó de forma consistente entre los diferentes escenarios metanalíticos y también lo observaron Sweeting y cols. (Sweeting 2004).

El hallazgo se observó a pesar de que el método sólo produjo una aproximación al *odds ratio*. Para efectos muy grandes (p.ej. cociente de riesgo = 0,2), cuando se conoce que la aproximación es deficiente, los efectos del tratamiento se subestimaron, pero el método de Peto tuvo el mejor rendimiento entre todos los métodos considerados para riesgos de eventos de 1 en 1000, y el sesgo nunca fue mayor del 6% del riesgo del grupo control.

En otras circunstancias (es decir, riesgos de eventos por encima del 1%, efectos muy grandes con riesgos de eventos alrededor del 1% y metanálisis en los que muchos estudios están significativamente desequilibrados), los métodos con mejor rendimiento fueron el OR de Mantel-Haenszel sin correcciones de celdas cero, la regresión logística y un método exacto. Ninguno de estos métodos está disponible en RevMan.

Los métodos que se deberían evitar con los eventos poco frecuentes son los métodos de la varianza inversa (incluido el método de efectos aleatorios de DerSimonian y Laird). Los mismos incorporan directamente la varianza del estudio en la estimación de su contribución al metanálisis, pero habitualmente se basan en una aproximación de la varianza de una muestra grande, la cual no se supone que se utilice cuando los eventos son poco frecuentes. El método de DerSimonian y Laird es el único método de efectos aleatorios generalmente disponible en el programa informático metanalítico. Es de señalar que la incorporación de la heterogeneidad en una estimación de un efecto del tratamiento sería una consideración secundaria cuando se intenta producir estimaciones de los efectos de datos escasos (la duda principal es discernir si hay alguna señal de un efecto en los datos).

16.10 Información del capítulo

Editores: Julian PT Higgins, Jonathan J Deeks y Douglas G Altman del Grupo Cochrane de Métodos Estadísticos.

La versión en inglés de este capítulo se debería citar como: Higgins JPT, Deeks JJ, Altman DG (editors). Chapter 16: Special topics in statistics. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Autores que contribuyeron: Doug Altman, Deborah Ashby, Ralf Bender, Catey Bunce, Marion Campbell, Mike Clarke, Jon Deeks, Simon Gates, Julian Higgins, Nathan Pace y Simon Thompson.

Agradecimientos: Se agradece particularmente a Joseph Beyene, Peter Gøtzsche, Steff Lewis, Georgia Salanti, Stephen Senn e Ian White por sus útiles comentarios sobre las primeras versiones. Para los detalles del Grupo Cochrane de Métodos Estadísticos, ver Capítulo 9 (Recuadro 9.8.a).

16.11 Referencias

Abrams 2005
Abrams KR, Gillies CL, Lambert PC. Meta-analysis of heterogeneously reported trials assessing change from baseline. Statistics in Medicine 2005; 24: 3823-3844.

Bauer 1991
Bauer P. Multiple testing in clinical trials. *Statistics in Medicine* 1991; 10: 871-889.

Bender 2001
Bender R, Lange S. Adjusting for multiple testing - when and how? *Journal of Clinical Epidemiology* 2001; 54: 343-349.

Bender 2008
Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL, Thorlund K. Dealing with multiplicity in systematic reviews. *Journal of Clinical Epidemiology* 2008; 54: 343-349.

Bradburn 2007
Bradburn MJ, Deeks JJ, Berlin JA, Russell LA. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine* 2007; 26: 53-77.

Bucher 1997
Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta- analysis of randomized controlled trials. *Journal of Clinical Epidemiology* 1997; 50: 683-691.

Caldwell 2005
Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; 331: 897-900.

Campbell 2000
Campbell M, Grimshaw J, Steen N. Sample size calculations for cluster randomised trials. Changing Professional Practice in Europe Group (EU BIOMED II Concerted Action). *Journal of Health Services Research and Policy* 2000; 5: 12-16.

Chan 2005
Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *The Lancet* 2005; 365: 1159-1162.

Chen 2005
Chen T, Hoppe FM. Simultaneous confidence intervals. In: Armitage P, Colton T (editors). *Encyclopedia of Biostatistics* (2nd edition). Chichester (UK): John Wiley & Sons, 2005.

Cook 2005
Cook RJ, Dunnett CW. Multiple comparisons. In: Armitage P, Colton T (editors). *Encyclopedia of Biostatistics* (2nd edition). Chichester (UK): John Wiley & Sons, 2005.

CPMP Working Party on Efficacy of Medicinal Products 1995
CPMP Working Party on Efficacy of Medicinal Products. Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products. *Statistics in Medicine* 1995; 14: 1659-1682.

Dmitrienko 2006
Dmitrienko A, Hsu JC. Multiple testing in clinical trials. In: Kotz S, Balakrishnan N, Read CB, Vidakovic B (editors). *Encyclopedia of Statistical Sciences* (2nd edition). Hoboken (NJ): John Wiley & Sons, 2006.

Donner 1980
Donner A, Koval JJ. The estimation of intraclass correlation in the analysis of family data. *Biometrics* 1980; 36: 19-25.

Donner 2000
Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. London (UK): Arnold, 2000.

Donner 2001
Donner A, Piaggio G, Villar J. Statistical methods for the meta-analysis of cluster randomized trials. *Statistical Methods in Medical Research* 2001; 10: 325-338.

Donner 2002
Donner A, Klar N. Issues in the meta-analysis of cluster randomized trials. *Statistics in Medicine* 2002; 21: 2971-2980.

Elbourne 2002
Elbourne DR, Altman DG, Higgins JPT, Curtin F, Worthington HV, Vaillancourt JM. Meta-analyses involving cross-over trials: methodological issues. *International Journal of Epidemiology* 2002; 31: 140-149.

Eldridge 2004
Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clinical Trials* 2004; 1: 80-90.

Farrin 2005
Farrin A, Russell I, Torgerson D, Underwood M, UK BEAM Trial Team. Differential recruitment in a cluster randomized trial in primary care: the experience of the UK back pain, exercise, active management and manipulation (UK BEAM) feasibility study. *Clinical Trials* 2005; 2: 119-124.

Follmann 1992
Follmann D, Elliott P, Suh I, Cutler J. Variance imputation for overviews of clinical trials with continuous response. *Journal of Clinical Epidemiology* 1992; 45: 769-773.

Freeman 1989
Freeman PR. The performance of the two-stage analysis of two-treatment, two-period cross-over trials. *Statistics in Medicine* 1989; 8: 1421-1432.

Furukawa 2006
Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. *Journal of Clinical Epidemiology* 2006; 59: 7-10.

Gamble 2005
Gamble C, Hollis S. Uncertainty method improved on best-worst case analysis in a binary meta-analysis. *Journal of Clinical Epidemiology* 2005; 58: 579-588.

Glenny 2005
Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, Bradburn M, Eastwood AJ. Indirect comparisons of competing interventions. *Health Technology Assessment* 2005; 9: 26.

Hahn 2005
Hahn S, Puffer S, Torgerson DJ, Watson J. Methodological bias in cluster randomised trials. *BMC Medical Research Methodology* 2005; 5: 10.

Hanley 1983
Hanley JA, Lippman-Hand A. If nothing goes wrong, is everything all right? Interpreting zero numerators. *JAMA* 1983; 249: 1743-1745.

Health Services Research Unit 2004
Health Services Research Unit. Database of ICCs: Spreadsheet (Empirical estimates of ICCs from changing professional practice studies) [page last modified 11 Aug 2004]. Available from: http://www.abdn.ac.uk/hsru/epp/cluster.shtml (accessed 1 January 2008).

Higgins 2001
Higgins JPT, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Statistics in Medicine* 2001; 20: 2219-2241.

Higgins 2002
Higgins JPT, Spiegelhalter DJ. Being sceptical about meta-analyses: a Bayesian perspective on magnesium trials in myocardial infarction. *International Journal of Epidemiology* 2002; 31: 96-104.

Higgins 2008
Higgins JPT, White IR, Wood AM. Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clinical Trials* 2008; 5: 225-239.

Hollis 1999
Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999; 319: 670-674.

Hollis 2002
Hollis S. A graphical sensitivity analysis for clinical trials with non-ignorable missing binary outcome. *Statistics in Medicine* 2002; 21: 3823-3834.

Juszczak 2003
Juszczak E, Altman D, Chan AW. A review of the methodology and reporting of multi-arm, parallel group, randomised clinical trials (RCTs). *3rd Joint Meeting of the International Society for Clinical Biostatistics and Society for Clinical Trials*, London (UK), 2003.

Khan 1996
Khan KS, Daya S, Collins JA, Walter SD. Empirical evidence of bias in infertility research: overestimation of treatment effect in crossover trials using pregnancy as the outcome measure. *Fertility and Sterility* 1996; 65: 939-945.

Koch 1996
Koch GG, Gansky SA. Statistical considerations for multiplicity in confirmatory protocols. *Drug Information Journal* 1996; 30: 523-534.

Lathyris 2007
Lathyris DN, Trikalinos TA, Ioannidis JP. Evidence from crossover trials: empirical evaluation and comparison against parallel arm trials. *International Journal of Epidemiology* 2007; 36: 422-430.

Lee 2005a
Lee LJ, Thompson SG. Clustering by health professional in individually randomised trials. *BMJ* 2005; 330: 142-144.

Lee 2005b
Lee SHH. *Use of the two-stage procedure for analysis of cross-over trials in four aspects of medical statistics* (PhD thesis). University of London, 2005.

Lewis 1993
Lewis JA, Machin D. Intention to treat--who should use ITT? *British Journal of Cancer* 1993; 68: 647-650.

Little 2004
Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd edition). Hoboken (NJ): John Wiley & Sons, 2004.

Lunn 2000
Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; 10: 325-337.

Marinho 2003
Marinho VCC, Higgins JPT, Logan S, Sheiham A. Fluoride toothpaste for preventing dental caries in children and adolescents. *Cochrane Database of Systematic Reviews* 2003, Issue 1. Art No: CD002278.

McAlister 2003
McAlister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. *JAMA* 2003; 289: 2545-2553.

Mills 2005
Mills EJ, Chan AW, Guyatt GH, Altman DG. Design, analysis, and presentation of cross-over trials. *5th Peer Review Congress*, Chicago (IL), 2005.

Murray 1995
Murray DM, Short B. Intraclass correlation among measures related to alcohol-use by young-adults - estimates, correlates and applications in intervention studies. *Journal of Studies on Alcohol* 1995; 56: 681-694.

Newcombe 2000
Newcombe RN, Altman DG. Proportions and their differences. In: Altman DG, Machin D, Bryant TN, Gardner MJ (editors). *Statistics with Confidence* (2nd edition). London (UK): BMJ Books, 2000.

Newell 1992
Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. *International Journal of Epidemiology* 1992; 21: 837-841.

Ottenbacher 1998
Ottenbacher KJ. Quantitative evaluation of multiplicity in epidemiology and public health research. *American Journal of Epidemiology* 1998; 147: 615-619.

Puffer 2003
Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ* 2003; 327: 785-789.

Qizilbash 1998
Qizilbash N, Whitehead A, Higgins J, Wilcock G, Schneider L, Farlow M. Cholinesterase inhibition for Alzheimer disease: a meta-analysis of the tacrine trials. *JAMA* 1998; 280: 1777-1782.

Rao 1992
Rao JNK, Scott AJ. A simple method for the analysis of clustered binary data. *Biometrics* 1992; 48: 577-585.

Salanti 2008
Salanti G, Higgins J, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Statistical Methods in Medical Research* 2008; 17: 279-301.

Senn 2002
Senn S. *Cross-over Trials in Clinical Research* (2nd edition). Chichester (UK): John Wiley & Sons, 2002.

Smith 1995
Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine* 1995; 14: 2685-2699.

Song 2003
Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003; 325: 472-475.

Spiegelhalter 2004
Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester (UK): John Wiley & Sons, 2004.

Stewart 1995
Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Statistics in Medicine* 1995; 14: 2057-2079.

Sutton 2000
Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for Meta-analysis in Medical Research*. Chichester (UK): John Wiley & Sons, 2000.

Sutton 2001
Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 2001; 10: 277-303.

Sweeting 2004
Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine* 2004; 23: 1351-1375.

te Velde 1998
te Velde ER, Cohlen BJ, Looman CW, Habbema JD. Crossover designs versus parallel studies in infertility research. *Fertility and Sterility* 1998; 69: 357-358.

Thompson 2001
Thompson SG, Turner RM, Warn DE. Multilevel models for meta-analysis, and their application to absolute risk differences. *Statistical Methods in Medical Research* 2001; 10: 375-392.

Turner 2000
Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* 2000; 19: 3417-3432.

Ukoumunne 1999
Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technology Assessment* 1999; 3: 5.

Unnebrink 2001
Unnebrink K, Windeler J. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. *Statistics in Medicine* 2001; 20: 3931-3946.

White 2005
White IR, Thomas J. Standardized mean differences in individually-randomized and cluster-randomized trials, with applications to meta-analysis. *Clinical Trials* 2005; 2: 141-151.

White 2007
White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomized controlled trials. *Clinical Trials* 2007; 4: 125-139.

White 2008a
White IR, Higgins JPT, Wood A. Allowing for uncertainty due to missing data in meta-analysis. Part 1: Two-stage methods. *Statistics in Medicine* 2008; 27: 711-727.

White 2008b
White IR, Welton N, Wood A, Ades AE, Higgins JPT. Allowing for uncertainty due to missing data in meta-analysis. Part 2: Hierarchical models. *Statistics in Medicine* 2008; 27: 728-745.

Whiting-O'Keefe 1984
Whiting-O'Keefe QE, Henke C, Simborg DW. Choosing the correct unit of analysis in medical care experiments. *Medical Care* 1984; 22: 1101-1114.

ÍNDICE
CAPÍTULO 17: RESULTADOS NOTIFICADOS POR EL PACIENTE

PUNTOS CLAVE

17.1 ¿QUÉ SON LOS RESULTADOS NOTIFICADOS POR EL PACIENTE?

Figura 17.1.a: Clasificación de los resultados de los ensayos clínicos con la ilustración de los más importantes para los pacientes

17.2 RESULTADOS NOTIFICADOS POR EL PACIENTE Y REVISIONES COCHRANE

17.3 ESTADO DE SALUD Y CALIDAD DE VIDA COMO RESULTADOS RNP

Tabla 17.3.a: Definiciones de términos seleccionados relacionados con la calidad de vida
Tabla 17.3.b: Una taxonomía del estado de salud y las medidas de calidad de vida de Patrick y Erickson (Patrick 1993).

17.4 ASPECTOS EN LA MEDICIÓN DE LOS RESULTADOS NOTIFICADOS POR EL PACIENTE

17.4.1 Validez de los instrumentos
17.4.2 Capacidad de un instrumento para medir el cambio

17.5 LOCALIZACIÓN Y SELECCIÓN DE LOS ESTUDIOS CON RESULTADOS NOTIFICADOS POR EL PACIENTE

17.6 EVALUACIÓN Y DESCRIPCIÓN DE LOS RESULTADOS NOTIFICADOS POR EL PACIENTE

Tabla 17.6.a: Una lista de verificación para describir y evaluar los RNP en los ensayos clínicos

17.7 COMPARABILIDAD DE DIFERENTES MEDIDAS DE RESULTADO NOTIFICADO POR EL PACIENTE

17.8 INTERPRETACIÓN DE LOS RESULTADOS

17.8.1 Resúmenes de un estudio enfocado en un solo resultado notificado por el paciente
17.8.2 Resúmenes de estudios que utilizan más de un resultado notificado por el paciente
17.8.3 Cuando los estudios no abordan resultados notificados por el paciente

17.9 INFORMACIÓN DEL CAPÍTULO

Recuadro 17.9.a: El Grupo Cochrane de Métodos de Resultados Notificados por el Paciente

17.10 REFERENCIAS

CAPÍTULO 17: RESULTADOS NOTIFICADOS POR EL PACIENTE

Autores: Donald L Patrick, Gordon H Guyatt y Catherine Acquadro del Grupo Cochrane de Métodos de Resultados Notificados por el Paciente.

Puntos clave

- Los resultados notificados por el paciente (RNP) son informes que provienen directamente de los pacientes acerca de cómo se sienten o su estado de funcionamiento en relación con la afección de salud y su tratamiento, sin la interpretación de los profesionales de la atención sanitaria ni de cualquier otra persona.
- Los RNP pueden estar relacionados con síntomas, signos, estado funcional, percepciones u otros aspectos como la conveniencia y la tolerabilidad.
- Los ítems que reflejan los conceptos incluidos en un cuestionario RNP se obtienen de la población objetivo; es esencial que el paciente participe en la creación del cuestionario para la validez del contenido.
- En el sitio web del Grupo de Métodos RNP (www.cochrane-pro-mg.org) se proporciona un glosario en el que se encuentran las definiciones de términos poco familiares para los revisores.
- Los RNP no sólo son importantes cuando no están disponibles otras medidas objetivas del resultado de la enfermedad, sino también para representar qué es más importante para los pacientes acerca de su afección y su tratamiento.
- Los RNP pueden ser continuos o categóricos. Hay técnicas disponibles para combinar ambos tipos de medidas.
- Los revisores que deciden incluir los RNP en sus ensayos pueden necesitar lecturas previas acerca de los mismos para asegurarse de que los comprenden, en particular su validez y su capacidad para detectar cambios.
- En este capítulo se proporciona una lista de verificación sobre aspectos relacionados con los RNP que los autores deberían considerar antes de incorporar los RNP a sus revisiones y a las tablas “Resumen de los hallazgos”.
- Si una revisión terminada no logra registrar los RNP cuando éstos se seleccionaron como resultados importantes en el protocolo de la revisión, entonces se debería resaltar en la revisión como una deficiencia en la investigación actual sobre la eficacia del tratamiento.

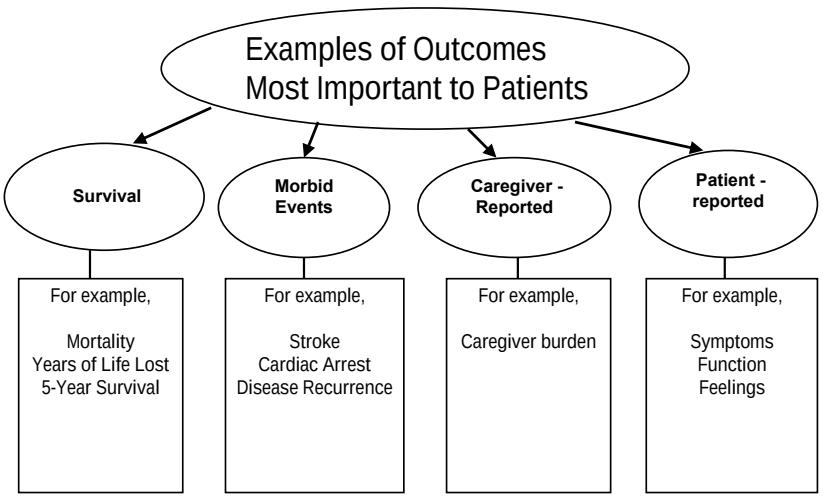
17.1 ¿Qué son los resultados notificados por el paciente?

Los resultados notificados por el paciente (RNP) son informes que provienen directamente de los pacientes acerca de su estado de funcionamiento o cómo ellos se sienten en relación con una afección de salud y su tratamiento, sin la interpretación de las respuestas de los pacientes por un médico ni por cualquier otra persona. Los RNP incluyen cualquier evaluación del tratamiento o de los resultados obtenida directamente de los pacientes mediante entrevistas, cuestionarios auto-cumplimentados, diarios u otras herramientas de recogida de datos como dispositivos manuales y formularios en la web (US Food and Drug Administration 2006). Los informes por poderes obtenidos de los cuidadores, profesionales de la salud, o padres o tutores (necesarios en algunas condiciones como cáncer avanzado o trastornos cognitivos) no se pueden considerar RNP y se deberían considerar una categoría de resultado separada.

Los RNP proporcionan la perspectiva del paciente sobre el beneficio del tratamiento; miden directamente los beneficios del tratamiento más allá de la supervivencia, la enfermedad y los marcadores fisiológicos; y con frecuencia son los resultados de mayor importancia para los pacientes. Los informes de los pacientes pueden incluir los signos y síntomas documentados en sus diarios, la evaluación de las sensaciones (clasificadas con más frecuencia como síntomas), los informes de conductas y habilidades (clasificadas con más frecuencia como estado funcional), percepciones o sentimientos generales de bienestar y otros informes que incluyen satisfacción con el tratamiento, calidad de vida general o relacionada con la salud y cumplimiento con el tratamiento. Los informes también pueden incluir efectos adversos o secundarios (ver Capítulo 14).

En algunas ocasiones los RNP se utilizan como resultados principales en los ensayos clínicos, particularmente cuando no están disponibles medidas alternativas de beneficio directo para conocer el bienestar de los pacientes. Con mayor frecuencia los RNP complementan los resultados primarios como la supervivencia, los indicadores de la enfermedad, las calificaciones de los profesionales sanitarios y las medidas fisiológicas o de laboratorio. La Figura 17.1.a muestra los resultados que con más frecuencia se consideran como importantes para los pacientes dentro de una clasificación de todos los resultados.

Figura 17.1.a: Clasificación de los resultados de los ensayos clínicos con la ilustración de los más importantes para los pacientes



Los RNP se pueden obtener mediante una medida (o instrumento) que sea específica para la enfermedad, específica para la afección, o genérica. Las medidas específicas para la

enfermedad describen gravedad, síntomas o limitaciones funcionales específicas para un estado determinado de enfermedad , afección o agrupamiento diagnóstico (p.ej. artritis o diabetes). Las medidas específicas para la afección describen los síntomas del paciente o las experiencias relacionadas con una afección o un problema específicos (p.ej. dolor lumbar) o relacionadas con intervenciones o tratamientos concretos (p.ej. cirugía de reemplazo de rodilla o de revascularización arterial coronaria). Las medidas genéricas están diseñadas para utilizarse con cualquier grupo de enfermedades o muestra poblacional.

En el sitio web del Grupo Cochrane de Métodos de Resultados Notificados por el Paciente (ver Recuadro 17.9.a) está disponible un glosario sobre los RNP.

17.2 Resultados notificados por el paciente y revisiones Cochrane

Los autores de revisiones sistemáticas seleccionarán los RNP para su inclusión según el alcance y los objetivos de su revisión. Los RNP son más importantes cuando resultados externamente observables e importantes para el paciente no están disponibles o son poco frecuentes. Para muchas afecciones, incluidos el dolor, los trastornos funcionales, la disfunción sexual y el insomnio no hay medidas biológicas satisfactorias disponibles. Las afecciones en las cuales los resultados sólo son conocidos por los propios pacientes, como la intensidad del dolor y las emociones, necesitan RNP como resultados primarios. Los RNP también son importantes cuando hay resultados observables disponibles, porque reflejan directamente lo que es importante para los pacientes.

Una parte inicial importante del proceso de revisión sistemática es definir y enumerar todos los resultados importantes para el paciente que son relevantes para el tema (Gyatt 2004) (ver Capítulo 5, Sección 5.4.1). Este paso está muy relacionado con la medición de los RNP. Muchos estudios primarios no logran medir aspectos de salud percibida y calidad de vida que son muy importantes para los pacientes. Cuando éste es el caso, la evidencia con respecto al impacto de las intervenciones sobre los RNP pueden ser mucho menos convincentes que la evidencia con respecto al impacto sobre indicadores de enfermedad como la morbilidad o la mortalidad. En caso extremo puede haber una línea en una tabla “Resumen de los hallazgos” que esté en blanco, que sea, por ejemplo, una línea que especifique la calidad de vida relacionada con la salud que esté en blanco porque el estudio no se enfocó directamente sobre este aspecto. La consideración previa cuidadosa de todos los resultados importantes para el paciente y la inclusión como una línea en blanco en una tabla “Resumen de los hallazgos” señalará lo que falta en cuanto a la medición del resultado en los ensayos aleatorizados elegibles y en otros estudios.

Es importante que los revisores comprendan la naturaleza de los RNP utilizados en los estudios incluidos en su revisión, y que comuniquen esta información al lector. En los ensayos clínicos los investigadores utilizan muchos instrumentos para obtener los RNP, y los métodos para desarrollar, validar y analizar los datos de RNP son diversos.

17.3 Estado de salud y calidad de vida como resultados RNP

Los resultados “estado de salud” y “calidad de vida” son una categoría importante de RNP. Con frecuencia los artículos publicados utilizan los términos “calidad de vida” (CdV), “estado funcional”, “calidad de vida relacionada con la salud” y “bienestar” con poca exactitud y de forma intercambiable, a pesar de las claras definiciones de los términos (ver Tabla 17.3.a).

Están disponibles diferentes tipos de instrumentos para medir el estado de salud y la calidad de vida (ver Tabla 17.3.b). Éstos pueden dar como resultado una puntuación general o un número indicador (que represente la repercusión de la intervención sobre la función física

o emocional, por ejemplo), un número índice (nuevamente una puntuación general, pero ponderada en términos de indicadores de muerte y salud total), un perfil (puntuaciones individuales de dimensiones o dominios) o una serie de pruebas (resultados múltiples que evalúan conceptos diferentes): ver Tabla 17.3.b.

La calidad de vida relacionada con la salud se puede medir mediante instrumentos genéricos o específicos, o una combinación de ambos. Si los investigadores están interesados en ir más allá de la enfermedad específica y posiblemente hacer comparaciones entre la repercusión de los tratamientos sobre la calidad de vida relacionada con la salud entre enfermedades o afecciones, deberían optar por medidas genéricas de calidad de vida relacionada con la salud que abarquen todas las áreas relevantes de la calidad de vida relacionada con la salud (incluidas, por ejemplo, movilidad, autocuidado, y función física, emocional y social) y se deberían diseñar para su administración a personas con cualquier tipo de problema de salud subyacente (o ningún problema). A estos instrumentos en ocasiones se les llama perfiles de salud; los perfiles de salud utilizados con más frecuencia son los formularios cortos de los instrumentos utilizados en el *Medical Outcomes Study* (Tarlov 1989, Ware 1995). De forma alternativa (o adicional) los ensayos aleatorizados y otros estudios se pueden haber basado en instrumentos específicos para la función (p.ej. sueño o función sexual), un problema (p.ej. dolor) o una enfermedad (p.ej. insuficiencia cardiaca, asma o síndrome de colon irritable).

La obtención de los conceptos e ítems para un cuestionario de RNP debería provenir de investigaciones cualitativas con pacientes, miembros de la familia, expertos médicos y de la bibliografía. Ver el Capítulo 20 para una orientación sobre cómo utilizar métodos cualitativos. La vinculación de los pacientes en el desarrollo de un cuestionario de RNP es esencial para asegurar la validez del contenido. Los conceptos que se incluyen y miden en un estudio incluido sólo se pueden determinar mediante el examen del contenido real de los ítems o aspectos incluidos en un instrumento que pretende medir la calidad de vida o la calidad de vida relacionada con la salud. El *concepto* es la “cosa” a medir. Los conceptos pueden estar relacionados con un ítem individual o con un subconjunto de ítems que se refieren al mismo concepto; con frecuencia se les llama dominios. Por ejemplo, un ítem que mida el dolor, una sensación conocida sólo por el paciente, sería un síntoma y el concepto de síntoma que se mide se puede etiquetar como dolor. Un ítem que evalúe la dificultad para subir escaleras sería un concepto relacionado con el funcionamiento físico y se podría etiquetar como subir escaleras o como parte del funcionamiento físico. La etiqueta de los conceptos varía ampliamente entre los investigadores y no existe un acuerdo acerca de la clasificación de los conceptos. Sin embargo, cada ítem, subdominio, dominio o la puntuación general aborda uno o más conceptos que los autores pueden identificar a partir del contenido, p.ej. el lenguaje, utilizado en la etiqueta para un ítem, un dominio o la puntuación general.

Los revisores pueden obtener mucho de lo que los autores de los estudios originales de desarrollo de RNP escriben acerca de la naturaleza o las fuentes de los ítems seleccionados para inclusión en un instrumento específico. Desafortunadamente los revisores a menudo tendrán que leer ellos mismos entre líneas los resultados de los ensayos clínicos publicados para probar y obtener una noción precisa de los conceptos o constructos que se consideran. Para lograr una comprensión completa, es posible que haya que hacer al menos una breve incursión en los artículos que describen el desarrollo y el uso anterior de los instrumentos de RNP incluidos en los estudios primarios.

Por ejemplo, los autores de una revisión Cochrane de terapia cognitiva conductual (TCC) para el tinitus incluyeron la calidad de vida como un resultado (Martínez-Devesa 2007). La calidad de vida se evaluó en cuatro ensayos mediante el *Tinnitus Handicap Questionnaire*, en un ensayo mediante el *Tinnitus Questionnaire* y en un ensayo mediante el *Tinnitus Reaction Questionnaire*. Las fuentes originales se citan en la revisión. Las citas de los artículos sobre las propiedades psicométricas también están disponibles en MEDLINE para los tres instrumentos y se pueden identificar fácilmente con una búsqueda que utilice el motor de búsqueda Google. La información sobre los ítems y los conceptos medidos están contenidos en estos artículos y los revisores pudieron comparar el contenido de los instrumentos.

Otro aspecto a considerar para comprender lo que se mide es cómo se ponderan los instrumentos de RNP. Muchos instrumentos específicos ponderan los ítems de forma similar cuando producen una puntuación general. Los instrumentos utilitarios diseñados principalmente para un análisis económico ponen un gran énfasis en la ponderación del ítem, e intentan presentar la calidad de vida relacionada con la salud como un proceso continuo ubicado entre la muerte y la salud total. Los lectores interesados en los aspectos que se han esbozado en el párrafo anterior pueden examinar un resumen antiguo pero todavía útil (Guyatt 1993).

Tabla 17.3.a: Definiciones de términos seleccionados relacionados con la calidad de vida

Término	Definición
Estado funcional	El rendimiento efectivo de un individuo o la capacidad para realizar funciones, tareas o actividades de importancia (p.ej. ir al trabajo, practicar deportes o mantener la casa).
Calidad de vida relacionada con la salud	Estado de salud personal. La calidad de vida relacionada con la salud se refiere habitualmente a los aspectos de la vida que están dominados o significativamente influidos por nuestro bienestar mental o físico.
Calidad de vida (CdV)	Una evaluación de todos los aspectos de la vida, incluido, por ejemplo, dónde vivimos, cómo vivimos y cómo actuamos. Incluye los factores de la vida como las circunstancias familiares, las finanzas, la casa y la satisfacción con el trabajo. (Ver también calidad de vida relacionada con la salud).
Bienestar	Estado físico y emocional subjetivo; cómo se siente un individuo; un estado mental diferente del funcionamiento, que se relaciona con las conductas y las actividades.

Tabla 17.3.b: Una taxonomía del estado de salud y las medidas de calidad de vida de Patrick y Erickson (Patrick 1993).

Medida	Fortalezas	Debilidades
Tipos de puntuaciones producidas		
Número indicador único.	Evaluación global; Útil para la población.	Puede ser difícil de interpretar.
Número índice único.	Representa el impacto neto; Útil para coste-efectividad.	En ocasiones no es posible separar la contribución de los dominios a la puntuación general.
Perfil de puntuaciones interrelacionadas.	Instrumento único; Es posible la contribución de los dominios a la puntuación general.	La extensión puede ser un problema; Es posible que no haya una puntuación general.
Batería de puntuaciones independientes.	Es posible un amplio rango de resultados relevantes.	No es posible correlacionar diferentes resultados con una escala de medición común; Puede necesitar ajuste para comparaciones múltiples; Puede necesitar la identificación de un resultado principal.
Rango de poblaciones y conceptos		
Genérico: aplicado a las enfermedades, afecciones, poblaciones y conceptos.	Ampliamente aplicable; Resume un rango de conceptos; Es posible la detección de efectos no anticipados.	Es posible que no responda al cambio; Es posible que no se centre en el interés del paciente; La extensión puede ser un problema; Los efectos pueden ser difíciles de interpretar.
Específico: aplicado a los individuos, enfermedades, afecciones, poblaciones o conceptos/dominios.	Más aceptables para las personas que responden; Puede responder mejor al cambio.	No es posible comparar afecciones o poblaciones; No es posible detectar efectos no anticipados.
Sistema de ponderación		
Utilidad: ponderaciones de preferencia de los pacientes, los proveedores o la comunidad.	Escala de intervalo; Se incorpora el criterio del paciente o del consumidor.	Puede ser difícil obtener las ponderaciones; Es posible que no difiera de una ponderación similar, la cual es más fácil de obtener.
Ponderación igual: los ítems se ponderan de manera similar o a partir de la frecuencia o la respuesta.	Técnicas más familiares; Parece ser más fácil de utilizar.	Puede estar influida por la prevalencia; No se pueden incorporar los intercambios.

17.4 Aspectos en la medición de los resultados notificados por el paciente

17.4.1 Validez de los instrumentos

La validez está relacionada con el hecho de si el instrumento mide lo que se supone que debería medir. La evidencia empírica de que los RNP miden los dominios de interés permiten realizar inferencias fuertes con respecto a la validez. Para proporcionar dichas pruebas los investigadores deberían asimilar las estrategias de validación de los psicólogos, quienes durante muchos años han intentado determinar si los cuestionarios que evalúan la inteligencia y las actitudes realmente miden lo que se supone que miden.

Las estrategias de validación incluyen:

- las relacionadas con el contenido: prueba de que los ítems y los dominios de un instrumento son apropiados y exhaustivos con respecto a su supuesta medición del/los concepto/s, la población y el uso;
- las relacionadas con la construcción: prueba de que las relaciones entre los ítems, los dominios y los conceptos forman una hipótesis a priori con respecto a las relaciones lógicas que deberían existir con otras medidas o características de los pacientes y los grupos de pacientes: y
- las relacionadas con el criterio (para un instrumento de RNP utilizado como herramienta diagnóstica): el grado en el cual las puntuaciones de un instrumento de RNP se relacionan con una medida criterio.

Establecer la validez incluye el examen de las relaciones lógicas que deberían existir entre las medidas de evaluación. Por ejemplo, es de esperar que los pacientes con una menor capacidad de ejercicio en la cinta de correr generalmente tengan una respiración más acortada en la vida diaria que los que tienen una mayor capacidad de ejercicio, y es de esperar que se observen correlaciones significativas entre una nueva medida de función emocional y los cuestionarios de función emocional existentes.

Cuando el interés se centra en evaluar el cambio en el tiempo, se examinan las correlaciones en las puntuaciones de cambio. Por ejemplo, los pacientes con deterioro en su capacidad de ejercicio en la cinta de correr deberían, en general, mostrar un aumento de la disnea, mientras que los que tienen una mejoría en su capacidad deberían presentar menos disnea. De manera similar, una nueva medida emocional debería mostrar una mejoría en los pacientes que mejoran en las medidas de función emocional existentes. El término técnico para este proceso es la prueba de la validez de la construcción del instrumento.

Los revisores deberían examinar y evaluar la evidencia de la validez de los RNP utilizados en los estudios incluidos. Desafortunadamente los ensayos aleatorizados y otros estudios que utilizan RNP pocas veces examinan la evidencia de la validez de los instrumentos que utilizan, pero los revisores pueden obtener cierta confianza de los planteamientos (respaldados por citas) de que los cuestionarios han sido validados anteriormente.

Una duda final acerca de la validez surge de si el instrumento de medición se utiliza con una población diferente o en un medio cultural y lingüísticamente diferente del medio en el que se desarrolló (típicamente, el uso de una versión que no sea en inglés de un cuestionario en idioma inglés). De manera ideal, se debería contar con pruebas de validez para la población reclutada en el ensayo aleatorizado. También de manera ideal, las medidas de RNP se deberían validar nuevamente en cada estudio mediante cualquier dato que esté disponible para la validación, por ejemplo, otras variables de resultado medidas. Los revisores deberían señalar, al evaluar la evidencia de validez, si la población evaluada en el ensayo es diferente a la utilizada en los estudios de validación.

17.4.2 Capacidad de un instrumento para medir el cambio

Cuando se utilizan instrumentos para evaluar los efectos del tratamiento, los mismos deberían ser capaces de medir las diferencias entre los grupos, si dichas diferencias existen en realidad. La asignación al azar debería asegurar que los participantes en los grupos de intervención experimental y control comiencen en el estudio con el mismo estado con respecto al concepto o la suposición que el RNP esté diseñado para medir. Los RNP deberían ser capaces de detectar lo que es importante para los pacientes y distinguir entre los participantes que siguen igual, mejoran o se deterioran en el transcurso del ensayo. Esto se denomina en ocasiones capacidad de respuesta, o sensibilidad al cambio.

Un instrumento con una capacidad deficiente para medir el cambio puede dar lugar a resultados falsos negativos en los cuales la intervención experimental mejora lo que siente el paciente, aunque el instrumento no pueda detectar la mejoría. Este problema puede ser particularmente importante para los cuestionarios genéricos que tienen la ventaja de cubrir todas las áreas relevantes de la calidad de vida relacionada con la salud, pero que tienen la desventaja de cubrir cada área de forma superficial. En los estudios que no muestran diferencias en los RNP entre la intervención experimental y control, un posible motivo es la falta de capacidad de respuesta del instrumento.

17.5 Localización y selección de los estudios con resultados notificados por el paciente

Los métodos de búsqueda para los RNP son los mismos que para otros resultados (ver Capítulo 6). Habitualmente se examinarán todos los informes recuperados por la estrategia de búsqueda de la revisión para identificar los que incluyan los RNP de interés. En ocasiones se puede utilizar una búsqueda separada adicional de RNP para complementar la estrategia estándar. Por ejemplo, si una revisión de ensayos aleatorizados y otros estudios en el área de asma no produce estudios que utilicen RNP, se puede realizar una búsqueda separada que incluya términos de búsqueda específicos para los RNP utilizados en el asma, como “calidad de vida específica para el asma”. Sin embargo, esto se basa en que se mencionen los RNP en el registro electrónico dentro de las bases de datos examinadas.

Los términos índice para los RNP difieren entre las principales bases de datos bibliográficas. Los revisores no deberían confiar en un único índice o subtítulo de término de búsqueda para identificar estudios que aborden RNP. Habitualmente se necesitan múltiples términos de búsqueda. Por ejemplo, Maciejewski y cols. utilizaron los siguientes términos índice de MEDLINE en su revisión sistemática para estimar el efecto de intervenciones para perder peso sobre la calidad de vida relacionada con la salud en ensayos aleatorizados (Maciejewski 2005): “Contingent valuation”; “Health status”; “Health-related Quality of Life”; “Psychological aspects”; “Psychosocial”; “Quality of life”; “Self-efficacy”; “SF-36”; “Utility”; “Well-being”; “Willingness to pay”. Las búsquedas de texto libre también deberían incluir tantos sinónimos relevantes como sea posible. La búsqueda debería combinar términos índice y términos de texto libre y es probable que se deba repetir en varias ocasiones.

Los revisores pueden encontrar útil el diseñar y utilizar una sección separada del formulario de obtención de datos utilizado en la revisión sistemática para incluir revisiones de métodos y resultados de RNP. Un ejemplo de este tipo de formulario se puede encontrar en el sitio web: www.cochrane-pro-mg.org/documents.html. Los revisores deberían estar atentos a formas alternativas de obtener los datos de los instrumentos, especialmente si se pueden obtener en formularios que faciliten el análisis de los datos en forma de variables continuas y resultados dicotómicos.

17.6 Evaluación y descripción de los resultados notificados por el paciente

La Tabla 17.6.a presenta aspectos seleccionados específicos para los RNP que los revisores deberían considerar cuando incorporan RNP en sus revisiones. Es posible que los autores deseen describir los RNP en detalle según esta lista de verificación en la tabla "Características de los estudios incluidos" o como una tabla adicional.

Tabla 17.6.a: Una lista de verificación para describir y evaluar los RNP en los ensayos clínicos

Basada en el Capítulo 7 de Patrick y Erickson, *Users' Guide to the Medical Literature*, las guías del CDC para la evaluación de los servicios de prevención comunitarios y los criterios utilizados por el *Medical Outcomes Trust* (Patrick 1993, Guyatt 1997, Zaza 2000, Lohr 2002).

1. ¿Qué midieron los RNP? a. ¿Qué conceptos midieron los RNP utilizados en el estudio? b. ¿Qué justificación (de haber alguna) proporcionaron los autores para la selección de los conceptos o constructos? c. ¿Los pacientes participaron en la selección de los resultados medidos por los RNP?
2. Omisiones a. ¿Se omitió algún aspecto importante de la salud (p.ej. síntomas, función, percepciones) o de la calidad de vida (p.ej. evaluación general, satisfacción con la vida) en este estudio a partir de las perspectivas del paciente, del médico, de otras personas importantes, contribuyentes u otros administradores y encargados de tomar decisiones?
3. Si los ensayos aleatorizados y otros estudios midieron RNP ¿cuáles fueron las estrategias de medición de los instrumentos? a. ¿Los investigadores utilizaron instrumentos que producen un indicador o un número índice únicos, un perfil o una batería de instrumentos? b. Si los investigadores midieron RNP, ¿utilizaron medidas específicas o genéricas, o ambas? c. ¿Quién exactamente completó los instrumentos?
4. ¿Los instrumentos funcionan de la forma que se supone que deberían funcionar (validez)? a. ¿Los instrumentos utilizados se han validado anteriormente (proporcionar la referencia)? ¿Se han presentado pruebas de la validación anterior para su uso en esta población? b. ¿Los instrumentos se validaron nuevamente en este estudio?
5. ¿Los instrumentos funcionan de la forma que se supone que deberían funcionar (capacidad para medir el cambio)? a. ¿Los RNP son capaces de detectar cambios en el estado del paciente, incluso si dichos cambios son pequeños?
6. ¿La magnitud del efecto (de haber alguno) puede ser comprensible para el lector? (¡Se debería intentar!) a. ¿Es posible proporcionar una estimación de la diferencia en los pacientes que logran un umbral de función o mejoría, y el número necesario a tratar (NNT) asociado?

17.7 Comparabilidad de diferentes medidas de resultado notificado por el paciente

Los investigadores pueden seleccionar diferentes instrumentos para medir los RNP, ya sea porque utilizan diferentes definiciones de un RNP concreto o porque seleccionan diferentes instrumentos para medir el mismo RNP. Por ejemplo, un investigador puede decidir utilizar un instrumento genérico para medir el estado funcional o un instrumento diferente específico para la enfermedad para medir este estado. La definición del resultado puede diferir o no. Los revisores deberían decidir cómo categorizar los RNP entre los estudios y cuándo combinar los resultados. Estas decisiones se basarán en las características del RNP, que será necesario extraer y describir en la revisión.

En muchas ocasiones los estudios que utilizan RNP harán mediciones iniciales y de seguimiento y el resultado de interés será, por lo tanto, la diferencia en el cambio a partir del valor inicial hasta el seguimiento entre los grupos de intervención y control. Entonces de manera ideal, para combinar los datos de dos RNP que están conceptualmente relacionados, habrá evidencia de correlaciones longitudinales fuertes del cambio en las dos medidas en los datos de pacientes individuales y evidencia de una capacidad de respuesta similar de los instrumentos. La evidencia adicional de apoyo puede provenir de las correlaciones de las diferencias entre el tratamiento y el control, o de la diferencia entre las mediciones de antes y después, entre los estudios. Si no es posible encontrar estos datos, se puede confiar en las correlaciones transversales en pacientes individuales en un punto temporal.

Por ejemplo, los dos instrumentos principales utilizados para medir la calidad de vida relacionada con la salud en pacientes con enfermedad obstructiva crónica son el *Chronic Respiratory Questionnaire* (CRQ) (Cuestionario Respiratorio Crónico) y el *St. George's Respiratory Questionnaire* (SGRQ) (Cuestionario Respiratorio de St. George). Las correlaciones entre los dos cuestionarios en estudios individuales han variado de 0,3 a 0,6 en las comparaciones transversales (correlaciones en un punto temporal) y en las longitudinales (correlaciones de cambio) (Rutten-van Mölken 1999, Singh 2001, Schünemann 2003, Schünemann 2005).

En una investigación posterior los investigadores examinaron las correlaciones entre la media de los cambios en el CRQ y en el SGRQ en 15 estudios que incluyeron 23 grupos de pacientes y encontraron una correlación de 0,88 (Puhan 2006). A pesar de esta correlación tan fuerte, el CRQ demostró tener una mayor capacidad de respuesta que el SGRQ: las medias estandarizadas de la respuesta del CRQ (mediana de las medias estandarizadas de la respuesta 0,51; IQR 0,19 a 0,98) fueron significativamente mayores ($p < 0,001$) que las asociadas con el SGRQ (mediana de las medias estandarizadas de la respuesta 0,26; IQR – 0,03 a 0,40). Es decir, en situaciones en las que ambos instrumentos se utilizaron juntos en el mismo estudio el CRQ produjo sistemáticamente efectos mayores del tratamiento. Como resultado, la combinación de los resultados de los ensayos que utilizaron estos dos instrumentos podría dar lugar a la subestimación del efecto del tratamiento en los estudios que utilicen el SGRQ.

Desafortunadamente, la mayoría de veces los datos detallados, como los descritos en el párrafo anterior, no estarán disponibles. Los investigadores tienen que confiar entonces en decisiones intuitivas acerca del grado en el cual diferentes instrumentos miden el mismo constructo subyacente. Por ejemplo, los autores de un metanálisis de intervenciones psicosociales en el tratamiento del síndrome premenstrual se enfrentaron a una gran cantidad de medidas de resultado, con 25 RNP descritos en sus nueve estudios elegibles. Para solucionar este problema dos investigadores examinaron de forma independiente cada instrumento (incluidos todos los dominios) y los agruparon en seis categorías conceptuales discretas; las discrepancias se resolvieron mediante discusión hasta lograr consenso. El análisis combinado de cada categoría incluyó entre dos y seis estudios.

Habitualmente los metanálisis de estudios que utilizan diferentes escalas de medición se realizarán mediante las diferencias de medias estandarizadas (DME; ver Capítulo 9, Sección 9.2.3). Sin embargo, las DME son muy problemáticas cuando el objetivo es comparar el cambio a partir del valor inicial en los grupos intervención y control, ya que las desviaciones estándar del cambio no miden la variación entre los pacientes (dependen también de la correlación entre las mediciones iniciales y finales; ver Capítulo 9, Sección 9.4.5.2).

Principios similares se aplican a los estudios en los cuales los revisores deciden centrarse en los datos disponibles que se presentan de forma dicotómica, o de los cuales los revisores pueden extraer datos de resultados dicotómicos con relativa facilidad. Por ejemplo, los investigadores que estudiaron el impacto de los flavanoides sobre los síntomas de las hemorroides encontraron que los ensayos aleatorizados elegibles no utilizaron de forma consistente medidas de síntomas similares; sin embargo, con excepción de uno de los 14 ensayos, todos registraron la proporción de pacientes sin síntomas, con mejoría de los síntomas, aún sintomáticos, o peor (Alonso-Coello 2006). En el análisis primario los investigadores consideraron los resultados de los pacientes sin síntomas y los pacientes con mejoría sintomática/alguna mejoría como equivalentes, y combinaron cada resultado de interés según una expectativa *a priori* de una magnitud y dirección similares del efecto del tratamiento.

Lo anterior hizo surgir la pregunta de cómo lidiar con los estudios que informan que los pacientes presentaron “alguna mejoría”. Los investigadores realizaron análisis en los que compararon el enfoque de dicotomizar, con la inclusión de “alguna mejoría” como un resultado positivo y como un resultado negativo (similar a ninguna mejoría). A menudo la dicotomización de los resultados es muy útil, particularmente para hacer los resultados fácilmente interpretables para los médicos y los pacientes. Las formas imaginativas, e incluso rigurosas de dicotomizar darán lugar a estadísticas resumen que proporcionan guías útiles para la práctica clínica.

El uso de múltiples instrumentos para medir un RNP concreto y la experimentación con múltiples métodos para el análisis puede dar lugar a la notificación selectiva de hallazgos más interesantes e introducir sesgos graves en una revisión sistemática. Los revisores que se centran en los RNP deberían estar alertos a este problema. Cuando sólo un escaso número de estudios elegibles haya descrito un determinado resultado , especialmente si es un resultado importante que es de esperar que los investigadores concientes midan, los revisores deberían señalar la posibilidad de sesgo de notificación (ver Capítulo 10).

17.8 Interpretación de los resultados

17.8.1 Resúmenes de un estudio enfocado en un solo resultado notificado por el paciente

Cuando un metanálisis incluye estudios que solamente describen un único RNP, presentado como una variable continua, un resultado combinado generará una diferencia de medias. El problema con esta diferencia de medias es que los médicos pueden tener dificultades con su interpretación. Por ejemplo, si se dice que la diferencia de medias entre la rehabilitación y la atención estándar en una serie de ensayos aleatorizados que utilizaron el *Chronic Respiratory Questionnaire* fue 1,0 (IC del 95%: 0,6; 1,5), muchos lectores no tendrán idea de si esto representa un efecto trivial pero importante, moderado o grande.

El autor de la revisión sistemática puede ayudar a la interpretación al señalar el rango de resultados posibles y el rango de resultados de la media de los grupos tratamiento y control en los estudios. Sin embargo, es de mayor utilidad (si está disponible) una estimación de la diferencia más pequeña que posiblemente los pacientes consideren importante (la diferencia mínima importante o DMI). Hay varios métodos para generar estimaciones de la DMI,

incluido el uso de calificaciones globales de cambio (Guyatt 2002). De manera ideal, los revisores presentarán las estimaciones de la DMI en el resumen. Por ejemplo, los investigadores que examinaron la repercusión de la rehabilitación respiratoria en pacientes con enfermedad pulmonar crónica sobre la calidad de vida relacionada con la salud, presentaron en su resumen que "para dos características importantes de la calidad de vida relacionada con la salud, la disnea y la destreza, el efecto general fue mayor que la DMIC: 1,0 (IC del 95%: 0,6; 1,5) y 0,8 (0,5; 1,2), respectivamente, comparada con una DMIC de 0,5" (Lacasse 1996).

Aunque lo anterior es muy útil, posiblemente tienta a los médicos a hacer inferencias inapropiadas. Si la DMI es 0,5 y la diferencia de medias entre los tratamientos es 0,4 los médicos pueden inferir que nadie se beneficia de la intervención. Si la diferencia de medias es 0,6 pueden concluir que todos se benefician. Ambas inferencias pueden ser erróneas. Primero, ignoran la incertidumbre (intervalos de confianza) alrededor de la estimación puntual. Más importante, ignoran la variación (desviación estándar) en las respuestas entre los individuos.

También es posible que los investigadores proporcionen una definición de "paciente que responde" para ayudar a interpretar los resultados (ver Capítulo 12, Sección 12.6.1). Es útil conocer la definición que caracteriza a un paciente individual como alguien que responde al tratamiento. Esta definición de paciente que responde se basa en criterios pre-especificados respaldados por pruebas empíricamente derivadas que apoyan la definición de paciente que responde como una medida de beneficio. Los métodos para definir a un paciente que responde incluyen: (1) un cambio pre-especificado a partir del valor inicial en una o más escalas; (2) un cambio de cierto tamaño o mayor en una puntuación (p.ej. un cambio de 2 puntos en una escala de 8 puntos); y (3) un cambio porcentual a partir del valor inicial.

17.8.2 Resúmenes de estudios que utilizan más de un resultado notificado por el paciente

Como en la discusión de la Sección 17.8.1 anteriormente señalada, cuando se combinan los RNP, la diferencia de medias deja de ser una posible medida de efecto, por lo que se reemplaza con la diferencia de medias estandarizadas (DME) (ver Capítulo 9, Sección 9.2.3). Desafortunadamente, no hay formas completamente satisfactorias para proporcionar un sentido de la magnitud del efecto en el RNP cuando hay que confiar en la DME para generar un resumen. Es posible ofrecer a los lectores reglas empíricas estándar para interpretar los tamaños de los efectos (por ejemplo, 0,2 representa un efecto pequeño; 0,5 un efecto moderado y 0,8 un efecto grande (Cohen 1988) o alguna variación (< 0,41 = pequeño; 0,40 a 0,70 = moderado; > 0,70 = grande). Otro enfoque, quizás menos satisfactorio, indica que una diferencia de medias estandarizada de 0,5 se aproxima, en muchos casos, a una diferencia mínima importante (Norman 2003).

Los métodos generales de descripción e interpretación de los RNP, así como otros resultados clínicos con respecto a la obtención de inferencias y conclusiones, se analizan en el Capítulo 12 (Sección 12.6)

17.8.3 Cuando los estudios no abordan resultados notificados por el paciente

Muchos estudios primarios no logran medir aspectos de la salud percibida ni la calidad de vida que son muy importantes para los pacientes. Cuando esto ocurre, la evidencia con respecto a la repercusión de las intervenciones sobre los RNP pueden ser mucho menos convincentes que la evidencia con respecto a la repercusión sobre indicadores de enfermedad como la morbilidad o la mortalidad. En sumo grado, el estudio puede no abordar directamente los RNP. La consideración previa exhaustiva de todos los resultados importantes para el paciente resaltaría los aspectos que faltan en la medición del resultado

en los ensayos aleatorizados elegibles y en otros estudios. Esta omisión se debería señalar en las conclusiones de los revisores como una implicación para investigaciones futuras.

17.9 Información del capítulo

Autores: Donald L Patrick, Gordon H Guyatt y Catherine Acquadro del Grupo Cochrane de Métodos de Resultados Notificados por el Paciente.

La versión en inglés de este capítulo se debería citar como: Patrick D, Guyatt GH, Acquadro C. Chapter 17: Patient-reported outcomes. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: Jason Busse, Peter Fayers, Toshi Furukawa, Madeleine King y Milo Puhan por proporcionar comentarios sobre los borradores.

Recuadro 17.9.a: El Grupo Cochrane de Métodos de Resultados Notificados por el Paciente

El objetivo principal del Grupo Cochrane de Métodos de Resultados Notificados por el Paciente (PRO MG) es asesorar a los revisores Cochrane acerca de cuándo y cómo incorporar datos del estado de salud y la calidad de vida a las revisiones sistemáticas. Algunos Grupos Cochrane de Revisión han tenido dificultades cuando incorporan datos de RNP en las revisiones. Ejemplos de estas dificultades incluyen la combinación y la interpretación de los datos y la evaluación de la validez de las escalas de RNP.

El PRO MG tiene como objetivos:

- refinar los métodos de búsqueda bibliográfica de estudios de RNP;
- desarrollar métodos para revisar sistemáticamente estudios de calidad de vida relacionada con la salud;
- refinar los métodos para el metanálisis de los estudios de RNP (en colaboración con el Grupo de Métodos Estadísticos);
- refinar los métodos para el uso de medidas de RNP en las evaluaciones económicas en colaboración con el Grupo Campbell-Cochrane de Métodos Económicos; y
- asesorar sobre el desarrollo de programas informáticos.

El grupo proporciona asesoría a solicitud al Grupo Directivo de la Colaboración Cochrane, organiza talleres sobre aspectos y métodos de resultados de salud notificados por el paciente, en respuesta a las necesidades de la Colaboración y prepara recomendaciones para este *Manual*. Los miembros del Grupo participarán en la preparación de las revisiones Cochrane y asesorarán a los revisores mediante materiales escritos y talleres de entrenamiento. Los miembros del grupo ayudarán a los revisores a desarrollar protocolos y revisiones en los que se haya decidido incluir RNP.

Sitio web: www.cochrane-pro-mg.org/

17.10 Referencias

Alonso-Coello 2006
Alonso-Coello P, Zhou Q, Martinez-Zapata MJ, Mills E, Heels-Ansdell D, Johanson JF, Guyatt G. Meta-analysis of flavonoids for the treatment of haemorrhoids. *British Journal of Surgery* 2006; 93: 909-920.

Cohen 1988
Cohen J. *Statistical Power Analysis in the Behavioral Sciences* (2nd edition). Hillsdale (NJ): Lawrence Erlbaum Associates, Inc., 1988.

Guyatt 1993
Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Annals of Internal Medicine* 1993; 118: 622-629.

Guyatt 1997
Guyatt GH, Naylor CD, Juniper E, Heyland DK, Jaeschke R, Cook DJ. Users' guides to the medical literature. XII. How to use articles about health-related quality of life. Evidence-Based Medicine Working Group. *JAMA* 1997; 277: 1232-1237.

Guyatt 2002
Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clinic Proceedings* 2002; 77: 371-383.

Guyatt 2004
Guyatt G, Montori V, Devereaux PJ, Schünemann H, Bhandari M. Patients at the center: in our practice, and in our use of language. *ACP Journal Club* 2004; 140: A11-A12.

Lacasse 1996
Lacasse Y, Wong E, Guyatt GH, King D, Cook DJ, Goldstein RS. Meta-analysis of respiratory rehabilitation in chronic obstructive pulmonary disease. *The Lancet* 1996; 348: 1115-1119.

Lohr 2002
Lohr K. Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of Life Research* 2002; 11: 193-205.

Maciejewski 2005
Maciejewski ML, Patrick DL, Williamson DF. A structured review of randomized controlled trials of weight loss showed little improvement in health-related quality of life. *Journal of Clinical Epidemiology* 2005; 58: 568-578.

Martinez-Devesa 2007
Martinez-Devesa P, Waddell A, Perera R, Theodoulou M. Cognitive behavioural therapy for tinnitus. *Cochrane Database of Systematic Reviews* 2007, Issue 1. Art No: CD005233.

Norman 2003
Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical Care* 2003; 41: 582-592.

Patrick 1993
Patrick DL, Erickson P. *Health Status and Health Policy: Quality of Life in Health Care Evaluation and Resource Allocation*. New York (NY): Oxford University Press, 1993.

Puhan 2006
Puhan M, Soesilo I, Guyatt GH, Schünemann HJ. Combining scores from different patient

reported outcome measures in meta-analyses: when is it justified? *Health and Quality of Life Outcomes* 2006; 4: 94.

Rutten-van Mölken 1999
Rutten-van Mölken M, Roos B, Van Noord JA. An empirical comparison of the St George's Respiratory Questionnaire (SGRQ) and the Chronic Respiratory Disease Questionnaire (CRQ) in a clinical trial setting. *Thorax* 1999; 54: 995-1003.

Schünemann 2003
Schünemann HJ, Griffith L, Jaeschke R, Goldstein R, Stubbing D, Guyatt GH. Evaluation of the minimal important difference for the feeling thermometer and the St. George's Respiratory Questionnaire in patients with chronic airflow obstruction. *Journal of Clinical Epidemiology* 2003; 56: 1170-1176.

Schünemann 2005
Schünemann HJ, Goldstein R, Mador MJ, McKim D, Stahl E, Puhan MA, Griffith LE, Grant B, Austin P, Collins R, Guyatt GH. A randomised trial to evaluate the self-administered standardised chronic respiratory questionnaire. *European Respiratory Journal* 2005; 25: 31-40.

Singh 2001
Singh SJ, Sodergren SC, Hyland ME, Williams J, Morgan MD. A comparison of three disease-specific and two generic health-status measures to evaluate the outcome of pulmonary rehabilitation in COPD. *Respiratory Medicine* 2001; 95: 71-77.

Tarlov 1989
Tarlov AR, Ware JE, Jr., Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study. An application of methods for monitoring the results of medical care. *JAMA* 1989; 262: 925-930.

US Food and Drug Administration 2006
US Food and Drug Administration. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims [February 2006]. Available from: http://www.fda.gov/cber/gdlns/prolbl.htm (accessed 1 January 2008).

Ware 1995
Ware JE, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Medical Care* 1995; 33: AS264-AS279.

Zaza 2000
Zaza S, Wright-De Agüero LK, Briss PA, Truman BI, Hopkins DP, Hennessy MH, Sosin DM, Anderson L, Carande-Kulis VG, Teutsch SM, Pappaioanou M, Task Force on Community Preventive Services. Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. *American Journal of Preventive Medicine* 2000; 18 (Suppl 1): 44-74.

ÍNDICE

CAPÍTULO 18: REVISIONES DE DATOS DE PACIENTES INDIVIDUALES

PUNTOS CLAVE

18.1 INTRODUCCIÓN

18.1.1 ¿Qué es una revisión de DPI?
18.1.2 ¿Cuándo se debería realizar una revisión de DPI?
Recuadro 18.1.a: Beneficios potenciales de los DPI
18.1.3 ¿En qué difieren los métodos de revisión de los DPI?
18.1.4 ¿Cómo se organizan las revisiones de DPI?
18.1.5 ¿Qué áreas de atención sanitaria utilizan el enfoque de DPI?
18.1.6 Si piensa realizar una revisión de DPI ¿qué debería hacer primero?

18.2 LA NATURALEZA COLABORATIVA DEL METANÁLISIS DE DPI

18.2.1 Grupos colaborativos
18.2.2 Negociación de colaboración
18.2.3 Confidencialidad

18.3 PROCESAMIENTO DE LOS DATOS

18.3.1 Decidir qué datos obtener
18.3.2 Formato de los datos
18.3.3 Recodificación y redefinición de las variables suministradas
18.3.4 Verificación de los datos suministrados
Introducción
18.3.4.1 Verificación de datos faltantes o duplicados
18.3.4.2 Verificación de la credibilidad
18.3.4.3 Verificación de la asignación al azar
18.3.4.4 Verificación de que la información es actualizada

18.4 ANÁLISIS

18.4.1 Ventajas del análisis
18.4.2 Enfoque general
18.4.3 Análisis de tiempo hasta el evento
18.4.4 Presentación de análisis actualizados: resultados a largo plazo
18.4.5 Análisis de subgrupos
18.4.6 Análisis adicionales
18.4.7 Programa informático

18.5 LIMITACIONES Y ADVERTENCIAS

18.5.1 ¿Qué no puede asegurar una revisión de DPI?
18.5.2 Estudios no disponibles
18.5.3 Decisión de cuándo es apropiada una revisión de DPI

18.6 INFORMACIÓN DEL CAPÍTULO

Recuadro 18.6.a: El Grupo Cochrane de Métodos de Metanálisis de Datos de Pacientes Individuales.

18.7 REFERENCIAS

CAPÍTULO 18: REVISIONES DE DATOS DE PACIENTES INDIVIDUALES

Autores: Lesley A Stewart, Jayne F Tierney y Mike Clarke del Grupo Cochrane de Métodos de Metanálisis de Datos de Pacientes Individuales.

Puntos clave

- En un metanálisis de datos de pacientes individuales (DPI) se solicitan directamente los datos originales de la investigación de cada participante en cada estudio a los investigadores responsables del estudio.
- El acceso a los datos "crudos" de cada estudio permite verificar los datos, hacer una exploración minuciosa y volver a analizar los datos de forma consistente.
- El metanálisis de DPI tiene beneficios concretos cuando la información publicada no permite realizar una revisión de buena calidad, o cuando se necesitan determinados tipos de análisis que no se pueden realizar con datos resumen.
- La mayoría de los metanálisis de DPI se realizan y publican por un grupo colaborativo, que incluye un equipo o una secretaría de proyecto, a los investigadores que contribuyen con los datos de sus estudios y con frecuencia también incluye un grupo asesor.
- Un enfoque de DPI habitualmente consume más tiempo y costos que una revisión sistemática convencional que cuenta con datos publicados o agregados.
- Puede haber circunstancias en las que los beneficios de obtener los DPI sean marginales; en otras puede ser vital.

18.1 Introducción

18.1.1 ¿Qué es una revisión de DPI?

Un metanálisis de datos de pacientes individuales (DPI) es un tipo específico de revisión sistemática. En lugar de extraer los datos de las publicaciones del estudio, los datos originales de investigación de cada participante en un estudio incluido se solicitan directamente a los investigadores responsables del estudio. Estos datos se pueden volver a analizar centralmente y, si es apropiado, combinarlos en los metanálisis. Las revisiones Cochrane se pueden realizar como revisiones de DPI, pero las revisiones de DPI habitualmente necesitan un equipo dedicado y pueden ser difíciles de realizar en el "tiempo libre". Este enfoque necesita habilidades determinadas y habitualmente consume más tiempo y costos que una revisión sistemática convencional que cuenta con los datos publicados o agregados. Sin embargo, las revisiones de DPI ofrecen beneficios relacionados especialmente con la calidad de los datos y los tipos de análisis que se pueden realizar (Stewart 1995, Stewart 2002). Por este motivo se consideran el "valor de referencia" de las revisiones sistemáticas. Este capítulo tiene como objetivo proporcionar una visión general del enfoque de los DPI en la revisión sistemática a fin de ayudar a los revisores a decidir si obtener los DPI puede ser útil y posible en su revisión. No se proporcionan detalles metodológicos y cualquiera que se proponga realizar su primera revisión de DPI debería solicitar asesoramiento apropiado y orientación de investigadores experimentados del Grupo de Métodos de Metanálisis de DPI (ver Recuadro 18.6.a).

18.1.2 ¿Cuándo se debería realizar una revisión de DPI?

Las revisiones de DPI se deberían considerar en las circunstancias en las que la información publicada no permite una revisión de buena calidad, o cuando se necesitan tipos de análisis determinados que no permiten utilizar enfoques estándar. Existen situaciones en las que no es posible el enfoque de DPI debido a que los datos se han destruido o perdido o cuando, a pesar de todos los esfuerzos, los investigadores no desean colaborar. También existen circunstancias en las que no es necesario, por ejemplo, si todos los datos requeridos están fácilmente disponibles en un formato adecuado en las publicaciones. En el Recuadro 18.1.a se proporcionan detalles adicionales de cuándo puede ser beneficioso obtener los DPI.

Recuadro 18.1.a: Beneficios potenciales de los DPI

Los DPI pueden ser beneficiosos en las situaciones siguientes.

- La mayoría de los estudios no están publicados o están publicados solamente en la literatura gris.
- Existe una descripción deficiente de los estudios (p.ej. la información presentada es inadecuada, selectiva o ambigua).
- Una alta proporción de individuos se ha excluido de los análisis publicados.
- La obtención de datos adicionales de resultados a largo plazo más allá de los publicados puede proporcionar hallazgos útiles (p.ej. para resultados de mortalidad o del desarrollo infantil).
- Medidas de resultado que se han definido de forma diferente entre los estudios.
- Se necesitan medidas de resultado de tiempo hasta el evento.
- Se necesitan análisis multivariados u otros análisis complejos.
- Es importante la exploración de interacciones entre las intervenciones y las características a nivel de paciente.

18.1.3 ¿En qué difieren los métodos de revisión de los DPI?

El enfoque general para el metanálisis de DPI es el mismo que para cualquier otra revisión sistemática, y los métodos utilizados sólo deberían diferir significativamente en las etapas de obtención de los datos, verificación y análisis. Como para cualquier revisión Cochrane se debería preparar un protocolo detallado, en el que se exponga el objetivo de la revisión, los aspectos específicos que se abordarán, los criterios de inclusión y exclusión de los estudios y los motivos por los que se solicitan los DPI, los métodos que se utilizarán y los análisis planificados. De forma similar, los métodos utilizados para identificar y revisar los estudios para la elegibilidad deberían ser los mismos, independientemente de si se solicitarán los DPI, aunque la implicación estrecha de los investigadores originales en el proyecto puede facilitar encontrar otros estudios realizados por ellos o por otros investigadores conocidos por ellos. El proyecto debería culminar con la preparación y difusión de un informe estructurado. Una revisión de DPI también puede incluir una reunión para presentar y discutir los resultados con los investigadores que colaboraron.

18.1.4 ¿Cómo se organizan las revisiones de DPI?

Las revisiones de DPI habitualmente se realizan como proyectos colaborativos en los cuales todos los investigadores contribuyen con información acerca de sus estudios en conjunto con los que dirigen el proyecto y forman parte de una colaboración activa. El proyecto es dirigido por un grupo o secretaría de proyecto local pequeño, que puede estar asistido en las tomas de decisiones importantes y estratégicas por un grupo de asesoramiento más grande. Habitualmente los resultados se publican en nombre del grupo colaborativo. La secretaría también puede ser responsable de organizar reuniones de los colaboradores para reunir a los investigadores en la discusión de los resultados preliminares.

18.1.5 ¿Qué áreas de atención sanitaria utilizan el enfoque de DPI?

Los metanálisis de DPI tienen antecedentes establecidos en la enfermedad cardiovascular y el cáncer, en los que la metodología se ha desarrollado ininterrumpidamente desde fines de los 80. En el cáncer, por ejemplo, actualmente existen más de 50 metanálisis de DPI de detección y tratamiento entre un amplio rango de sitios tumorales sólidos y neoplasias hematológicas (Clarke 1998). Los DPI también se utilizan en revisiones sistemáticas de muchos otros campos (Simmonds 2005), que incluyen infección por el VIH, demencia, epilepsia, depresión, paludismo, hernia y asma. El sitio web del Grupo de Métodos de Metanálisis de Datos de Pacientes Individuales de la Colaboración Cochrane incluye una base de datos de las revisiones de DPI en curso y finalizadas en las que se puede encontrar información adicional (ver Recuadro 18.6.a).

18.1.6 Si piensa realizar una revisión de DPI ¿qué debería hacer primero?

Antes de emprender una revisión de DPI se deberían considerar cuidadosamente las habilidades y el financiamiento requeridos para el éxito del proyecto y se debería solicitar entrenamiento y asesoramiento. El Grupo de Métodos de Metanálisis de Datos de Pacientes Individuales de la Colaboración Cochrane es un buen primer punto de contacto (Recuadro 18.6.a).

18.2 La naturaleza colaborativa del metanálisis de DPI

18.2.1 Grupos colaborativos

La mayoría de los metanálisis se realizan y publican por grupos colaborativos. Estos grupos comprenden el equipo o secretaría del proyecto que dirige la revisión de DPI, los miembros del grupo asesor (si existe uno) y los investigadores que contribuyen con los datos de sus estudios para volver a realizar el análisis.

18.2.2 Negociación de colaboración

Establecer la colaboración consume tiempo y esfuerzos considerables. Puede ser difícil rastrear a las personas responsables de los estudios elegibles y éstas inicialmente puedan rechazar participar en el metanálisis. Frecuentemente el primer enfoque puede ser mediante una carta con una invitación a colaborar en la que se explique el proyecto, se describa qué participación se entablará y cómo se dirigirá y publicará el metanálisis. Habitualmente la carta la envía el equipo del proyecto y se puede enviar a nombre del grupo asesor de la revisión. Generalmente en esta etapa se proporciona un protocolo para brindar información adicional, pero habitualmente en la primera correspondencia no se solicitan datos. También puede ser necesario establecer un contacto separado con el centro de datos o la organización de la investigación que son (o han sido) responsables del procesamiento de los datos del estudio, y a quienes será necesario enviarles consultas sobre los datos. Para estimular a los investigadores originales a tomar parte de la revisión de DPI es importante brindar tanto apoyo y ser tan flexible como sea posible, invertir el tiempo necesario para establecer relaciones y mantener a todos los colaboradores involucrados e informados sobre el progreso. Informes regulares y correos electrónicos

actualizados pueden ser formas útiles de mantener al grupo colaborativo actualizado e involucrado, especialmente si el proyecto se extiende por un período prolongado.

18.2.3 Confidencialidad

Naturalmente los investigadores necesitan proteger el uso de los datos de sus estudios y es conveniente asegurar que serán guardados de forma segura y utilizados apropiadamente. Por este motivo, con frecuencia se utiliza un acuerdo de confidencialidad firmado como un "contrato" entre los investigadores originales y el equipo de la revisión de DPI. Los detalles de tales acuerdos variarán, pero la mayoría establecerá que los datos serán guardados de forma segura, sólo tendrán acceso a ellos los miembros autorizados del equipo del proyecto y no serán copiados ni distribuidos en otra parte. Además la buena práctica insta a que los participantes individuales no sean identificados en los datos suministrados, por lo que los individuos son identificados solamente por un código identificador del estudio y no por su nombre. Éste parece ser un requerimiento cada vez más frecuente para obtener los DPI en algunos países en los que la legislación de protección de los datos exige que el participante no sea identificado a partir de los datos proporcionados. El envío de los datos por correo electrónico debería ser codificado, de ser posible.

18.3 Procesamiento de los datos

18.3.1 Decidir qué datos obtener

El protocolo debería especificar qué resultados y características del paciente se analizarán. Sin embargo, antes de emprender la obtención de los datos es sensato preguntarles a los investigadores originales qué datos están actualmente disponibles. Cuando se decida qué variables extraer, muchas veces es razonable comenzar por considerar cuidadosamente qué análisis se planifican y qué datos se necesitan para ellos. Esto minimiza la posibilidad de que no se solicite información esencial o que se obtengan datos innecesarios. Naturalmente, los investigadores se pueden molestar o ponerse suspicaces si se toman la molestia de proporcionar datos que no se analizan y publican posteriormente.

Aunque en muchos casos será posible obtener variables específicas para resultados y características como se definieron en los estudios individuales, puede que sea necesario considerar si hay detalles de los datos para los cuales se necesitan variables adicionales o constitutivas. Por ejemplo, si los estudios utilizan definiciones de resultados diferentes puede ser conveniente redefinir éstas para cada paciente de forma consistente entre los estudios y puede que se necesiten variables adicionales. Por ejemplo, para redefinir preeclampsia, se pudiera necesitar obtener los datos sobre la presión arterial sistólica y diastólica y la proteinuria.

18.3.2 Formato de los datos

Una vez que los investigadores originales están de acuerdo en colaborar, el siguiente paso es proporcionar instrucciones claras sobre qué datos necesitan proporcionar y en qué formato preferiblemente. El equipo del proyecto debería estar preparado para aceptar los datos en cualquier formato que sea más conveniente para quienes los suministran, ya sea

de forma electrónica, impresos o en formularios de papel, y debería estar preparado para recodificar la información cuando sea necesario. Sin embargo, aunque los primeros metanálisis de DPI en los 80 confiaban fundamentalmente en los datos proporcionados en papel, actualmente la mayoría de la información se suministra por correo electrónico o en discos, y a menudo los investigadores necesitan transformar o codificar sus datos según el formato especificado.

18.3.3 Recodificación y redefinición de las variables suministradas

Obtener los datos al nivel de participantes individuales permite la conversión entre diferentes estadiajes, escalas, categorías u otros sistemas de calificación y puede, por lo tanto, permitir la combinación de los datos de los estudios que de otra forma no sería posible debido a las diferencias entre las herramientas de obtención de los datos. Para permitir esto es importante que se soliciten los datos apropiados (ver Sección 18.3.1) y que los datos suministrados se recodifiquen o se transformen para reflejar definiciones comunes. Por ejemplo, si el resultado de interés es la preeclampsia, se necesitaría obtener los datos sobre la presión arterial y la proteinuria y considerarlos juntos para definir si se ha observado la preeclampsia (según la definición del protocolo de la revisión).

18.3.4 Verificación de los datos suministrados

Introducción

El objetivo de verificar los datos es aumentar la probabilidad de que los datos suministrados sean exactos, confirmar que los ensayos son adecuadamente aleatorizados y, de ser apropiado, asegurarse hasta donde sea posible de que los datos están actualizados. Los procedimientos exactos de verificación que se pueden realizar dependerán del área de asistencia sanitaria y de la cuestión abordada, así como de la naturaleza de los datos suministrados, pero hay cuatro áreas principales típicas:

18.3.4.1 Verificación de datos faltantes o duplicados

Cuando se reciben los datos es importante verificarlos tan pronto como sea posible para asegurar que se puedan leer e introducir en el sistema central de análisis. Por ejemplo, si los datos se reciben como adjuntos por correo electrónico, se debería verificar que los archivos se pueden abrir y que la información es del estudio correcto. En esta etapa es útil confirmar que se han recibido los datos de todos los individuos adecuados (habitualmente todos los asignados al azar), verificar que los números suministrados son consistentes con cualquier publicación u otra información y que, por ejemplo, no existen omisiones obvias ni duplicados en la secuencia del registro de pacientes ni en los números identificadores del estudio.

18.3.4.2 Verificación de la credibilidad

Las verificaciones de la credibilidad deberían incluir las verificaciones de los rangos de las variables, solicitarles a los investigadores originales que confirmen cualquier valor atípico extremo o inusual, por ejemplo, confirmar que los registros de pacientes insólitamente jóvenes o viejos, o de aquellos con niveles de colesterol anormalmente altos o bajos son en efecto correctos. La información suministrada también se debería contrastar con cualquier publicación relevante del estudio, por ejemplo, para confirmar que la distribución

de las características iniciales, el número de participantes y los resultados son consistentes (se debería tener en mente que el reclutamiento continuado o el seguimiento adicional pudieran alterar la información posterior de la publicación).

18.3.4.3 Verificación de la asignación al azar

Con frecuencia es útil verificar que la asignación al azar parece haber sido realizada de forma apropiada. Cuando están disponibles los datos de la asignación al azar, éstos se pueden explorar mediante el examen de los gráficos de acumulación en el tiempo; es de esperar que los números reclutados para cada intervención sean similares y que las curvas de reclutamiento se crucen con frecuencia. También puede ser informativo examinar la distribución de las asignaciones al azar según el día de la semana. Siempre que se haya asignado al azar un número razonable de individuos, es de esperar observar aproximadamente los mismos números asignados al azar a cada intervención en un día determinado de la semana, y que en los ensayos en los que la asignación al azar se realiza durante las horas normales de consulta, haya pocos, si es que hay algunos, participantes reclutados en días no esperados. También es útil verificar que los grupos de intervención estén equilibrados con respecto a características iniciales importantes y dentro de subgrupos importantes de participantes, pero hay que recordar que pueden aparecer por azar desequilibrios estadísticamente significativos.

18.3.4.4 Verificación de que la información es actualizada

Para los resultados en los cuales los eventos se observan durante un período prolongado, por ejemplo, la supervivencia en ensayos de cáncer, es importante verificar que el seguimiento esté tan actualizado como sea posible y que es consistente para cada uno de los grupos de intervención. La elaboración de una curva “inversa” de Kaplan Meier, basada solamente en los pacientes que no han presentado el evento de interés y en los que los datos censurados se utilicen como evento, puede proporcionar una verificación útil sobre el equilibrio del seguimiento entre los grupos.

Para cualquier estudio individual, los resultados de todas estas verificaciones se deberían considerar juntas para formarse una imagen general del estudio, de la calidad de los datos suministrados y de cualquier problema potencial. Cualquier duda se debería someter diplomáticamente a la consideración de los investigadores responsables. Habitualmente, los problemas se convierten en simples errores o malentendidos que pueden resolverse mediante discusión. Son poco frecuentes los problemas importantes que no se pueden resolver.

Se debería archivar una copia de los datos tal como se suministraron antes de realizar conversiones o modificaciones a los datos. Durante los procesos de verificación de los datos es importante que cualquiera de los cambios y alteraciones realizados a los datos suministrados se registre de forma adecuada.

18.4 Análisis

18.4.1 Ventajas del análisis

Tener acceso a los datos "crudos" de cada estudio permite la verificación, la exploración exhaustiva y el reanálisis de los datos de una forma consistente. Por lo tanto, no se debería confiar en la interpretación de la información ni en los análisis presentados en las publicaciones, no hay que estar limitado por los datos resumen proporcionados en formato tabular ni estar obligado a considerar la combinación de las estadísticas resumen de los estudios calculadas de maneras diferentes. Esto también evita problemas con los análisis originales; por ejemplo, se pueden realizar análisis según el principio de intención de tratar, incluso aunque en los análisis de los ensayos originales no se haya hecho así.

18.4.2 Enfoque general

La mayoría de los metanálisis de DPI hasta la fecha han utilizado un enfoque de dos etapas para el análisis. En la primera etapa cada estudio individual se analiza de la misma forma, como se estableció en el protocolo del metanálisis o en el plan de análisis. En el segundo paso, los resultados o las estadísticas resumen de cada uno de los análisis individuales del estudio se combinan para proporcionar una estimación combinada del efecto de la misma forma que para una revisión sistemática convencional (Simmonds 2005). Los enfoques más complejos utilizan modelados multinivel descritos para los datos binarios (Turner 2000), los datos continuos (Higgins 2001), los datos ordinales (Whitehead 2001) y los datos de tiempo hasta el evento (Tudor Smith 2005b), pero actualmente su aplicación es menos frecuente. Cuando no existe heterogeneidad entre los ensayos puede ser mejor evitar un enfoque de dos etapas del rango logarítmico estratificado para los datos de tiempo hasta el evento para una estimación de los efectos de la intervención más grandes (Tudor Smith 2005a).

18.4.3 Análisis de tiempo hasta el evento

La obtención de los DPI que incluyen el intervalo de tiempo entre la asignación al azar y el evento de interés permite que se realicen análisis de tiempo hasta el evento. Estos incluyen, por ejemplo, tiempo hasta la recuperación, tiempo sin convulsiones, tiempo hasta la concepción y tiempo hasta la muerte. De hecho, uno de los motivos principales por el que de los metanálisis de DPI han sido importantes en el campo del cáncer es que el análisis del tiempo hasta el evento de supervivencia es vital en la evaluación de los tratamientos. Es más probable que la mayoría de las intervenciones provoquen una prolongación de la supervivencia en lugar de una cura. Por lo tanto, es importante medir no sólo si ocurren muertes si no también el momento en el cual tienen lugar. Para poder realizar este tipo de análisis se necesita conocer el tiempo durante el cual el individuo permanece "sin eventos", lo que se logra habitualmente al obtener la fecha de asignación al azar, el estado del evento (es decir, si se observó el evento o no) y la fecha de la última evaluación para el evento. En ocasiones se obtendrá como el intervalo en días entre la asignación al azar y la evaluación más reciente para el evento. Los análisis de tiempo hasta el evento se realizan para cada ensayo para calcular los cocientes de riesgos instantáneos, los cuales luego se combinan en el metanálisis (ver Sección 9.4.9).

18.4.4 Presentación de análisis actualizados: resultados a largo plazo

Para resultados como la supervivencia, en los que los eventos pueden ocurrir de forma continua en el tiempo, los metanálisis de DPI pueden proporcionar una importante oportunidad para examinar los efectos de las intervenciones durante un período de tiempo prolongado. También pueden proporcionar una oportunidad para que los investigadores suministren datos más actualizados para resultados relevantes como la mortalidad que se puedan presentar para su estudio.

18.4.5 Análisis de subgrupos

La obtención de DPI también es la forma más práctica de realizar los análisis para investigar si cualquier efecto observado de una intervención es consistente entre tipos de participantes bien definidos, por ejemplo, si las mujeres obtienen un beneficio del tratamiento menor o mayor que los hombres. Con frecuencia en los análisis convencionales que utilizan datos agregados a partir de las publicaciones es muy difícil extraer datos compatibles suficientes para realizar análisis de subgrupos significativos, y especialmente difícil caracterizar a los individuos para el análisis de subgrupos (estratificado por estudio) definido por factores únicos o múltiples. La obtención de los DPI también permitirá realizar análisis más complejos como los modelados multinivel, para explorar las asociaciones entre los efectos de la intervención y las características de los pacientes.

18.4.6 Análisis adicionales

El acceso a los DPI también permite una exploración en profundidad de las características propias de los pacientes, independientemente de la intervención. Por ejemplo, la obtención de un gran conjunto de datos se puede utilizar en la construcción de índices pronósticos que pueden permitir predecir el resultado según las características del paciente (International Germ Cell Cancer Collaborative Group 1997).

18.4.7 Programa informático

Los DPI no se pueden analizar directamente en RevMan. Los datos se deberían analizar primero fuera de este programa informático, y las estadísticas resumen para cada estudio se deberían introducir en RevMan si se utiliza un enfoque de dos etapas. Para los resultados dicotómicos y continuos los datos se pueden introducir de la manera habitual. Para los resultados de tiempo hasta el evento, el número de eventos observados menos el número de eventos esperados y la varianza se pueden introducir mediante la opción “O – E y Varianza”. Alternativamente la opción varianza inversa se puede utilizar para analizar las estimaciones del efecto como los cocientes de riesgos instantáneos, las proporciones de la tasa o las estimaciones ajustadas.

Aunque muchos paquetes estadísticos estándar pueden realizar los análisis necesarios de los DPI provenientes de los estudios individuales, puede ser engorroso y demorado analizar de uno en uno cada resultado en cada estudio, y actualmente no está comercialmente disponible un programa informático con el que se pueda realizar el análisis directo, la combinación y los gráficos de los DPI en el metanálisis. El paquete de análisis no comercial “SCHARP” que analiza cada estudio, combina los resultados y

produce resultados tabulados y gráficos de bosque para los DPI dicotómicos, continuos y de tiempo hasta el evento, está disponible gratis para organizaciones sin fines de lucro. Este paquete basado en SAS ha sido desarrollado por el Grupo de Metanálisis de la Unidad de Ensayos Clínicos del Consejo de Investigaciones Médicas del Reino Unido. Éste está disponible contactando el Grupo de Métodos de Metanálisis de DPI (ver Recuadro 18.6.a).

18.5 Limitaciones y advertencias

18.5.1 ¿Qué no puede asegurar una revisión de DPI?

Aunque el enfoque de DPI puede ayudar a evitar los problemas asociados con los análisis y la publicación de los estudios, generalmente, no puede ayudar a evitar los sesgos asociados con el diseño o la realización del estudio. De existir tales problemas (los que también se pueden reflejar en las publicaciones del estudio y en cualquier revisión sistemática basada en las mismas), puede ser necesario excluir el estudio del metanálisis.

18.5.2 Estudios no disponibles

Con frecuencia la obtención de los DPI permite la inclusión de estudios que no es posible incluir en una revisión sistemática estándar debido a que no han sido publicados o no proporcionan información suficiente para permitir que se incluyan en los análisis. Esto puede ayudar a evitar muchos tipos de sesgo de publicación (Stewart 2002). Sin embargo, se debería asegurar que al limitar los análisis a los estudios que pueden proporcionar DPI no se introduzca sesgo a través de la disponibilidad selectiva de los datos de los estudios.

El éxito y la validez del enfoque de DPI requieren que los datos de todos o casi todos los estudios estén disponibles. Si la falta de disponibilidad está relacionada con los resultados del estudio, por ejemplo, si los investigadores están dispuestos a suministrar datos de los estudios con resultados prometedores pero vacilan en proporcionar datos de los que son menos favorables, entonces ignorar los estudios no disponibles puede sesgar los resultados de la revisión de DPI. Si se ha obtenido una gran proporción de datos, quizás del 90% o más de los individuos asignados al azar, se puede estar relativamente seguro de los resultados. Sin embargo, si la cantidad de información es menor es necesario ser convenientemente prudente al extraer conclusiones. El análisis de sensibilidad que combina los resultados de cualquier estudio no disponible (como se extrajo de las publicaciones o se obtuvo en un formato tabular) y la comparación de éstos con los resultados principales de los DPI es una ayuda útil para interpretar los datos. Las publicaciones de las revisiones de DPI que no pueden obtener DPI a partir de los estudios deberían señalar los motivos por los que los DPI no están disponibles y la probabilidad de sesgo resultante.

Al igual que para otras revisiones Cochrane, los metanálisis de DPI deberían señalar claramente qué estudios no se incluyeron y los motivos. Si sólo un número limitado de estudios puede proporcionar DPI para el análisis, entonces el valor del enfoque es cuestionable. Las experiencias en el cáncer han sido buenas y en la mayoría de los casos la perseverancia ha permitido que los datos estén disponibles para una alta proporción de ensayos elegibles. Esto hace que sea especialmente importante explorar la capacidad y la

buena voluntad de los investigadores principales para suministrar los DPI en etapas tempranas del proyecto.

18.5.3 Decisión de cuándo es apropiada una revisión de DPI

Cuando se comienza cualquier revisión sistemática es útil considerar cuidadosamente al inicio qué enfoque y qué tipo de datos serán los más apropiados. Se debería prestar atención concreta a los factores que es probable que introduzcan sesgo en la revisión. Puede haber casos en los que los beneficios de obtener los DPI sean apenas marginales y otros en los que pueden ser vitales.

18.6 Información del capítulo

Autores: Lesley A Stewart, Jayne F Tierney y Mike Clarke del Grupo Cochrane de Métodos de Metanálisis de Datos de Pacientes Individuales.

La versión en inglés de este capítulo se debería citar como: Stewart LA, Tierney JF, Clarke M. Chapter 19: Reviews of individual patient data. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: Se agradece a Paula Williamson sus útiles comentarios sobre la primera versión.

Recuadro 18.6.a: El Grupo Cochrane de Métodos de Metanálisis de Datos de Pacientes Individuales.

El Grupo Cochrane de Métodos de Metanálisis de Datos de Pacientes Individuales (GMMADPI) incluye individuos que participan o están interesados en la realización de revisiones sistemáticas que incluyen DPI y la investigación metodológica relacionada. El Grupo tiene como objetivo proporcionar asesoramiento a quienes realicen un metanálisis de DPI dentro de las revisiones Cochrane.

Las actividades de los miembros del GMMADPI incluyen las siguientes:

- Realizar metanálisis de DPI.
- Realizar investigaciones empíricas, por ejemplo, sobre los beneficios relativos de los metanálisis de DPI comparados con otras formas de revisiones sistemáticas, y utilizar la información obtenida por los metanálisis de DPI para explorar si los aspectos del diseño, el análisis y la publicación de los ensayos aleatorizados y las revisiones sistemáticas pueden ser fuentes de sesgo y heterogeneidad.
- Ayudar a los revisores Cochrane a decidir si sería apropiado para su revisión sistemática que la misma se realice mediante los DPI y, de ser así, ofrecer asesoramiento sobre cómo hacerla.
- Ofrecer talleres de formación en los Coloquios Cochrane y difundir materiales de formación de los mismos.
- Mantener un registro de revisiones que hayan utilizado (o utilizarán) DPI y una base de datos de proyectos y metanálisis de investigación metodológica.

Sitio web: www.ctu.mrc.ac.uk/cochrane/ipdmg

18.7 Referencias

Clarke 1998
Clarke M, Stewart L, Pignon JP, Bijnens L. Individual patient data meta-analysis in cancer. *British Journal of Cancer* 1998; 77: 2036-2044.

Higgins 2001
Higgins JPT, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Statistics in Medicine* 2001; 20: 2219-2241.

International Germ Cell Cancer Collaborative Group 1997
International Germ Cell Cancer Collaborative Group. International Germ Cell Consensus Classification: a prognostic factor-based staging system for metastatic germ cell cancers. *Journal of Clinical Oncology* 1997; 15: 594-603.

Simmonds 2005
Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials* 2005; 2: 209-217.

Stewart 1995
Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Statistics in Medicine* 1995; 14: 2057-2079.

Stewart 2002
Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation and the Health Professions* 2002; 25: 76-97.

Tudor Smith 2005a
Tudor Smith C, Williamson PR. Meta-analysis of individual patient data with time to event outcomes. *International Conference of the Royal Statistical Society*, Cardiff (UK), 2005.

Tudor Smith 2005b
Tudor Smith C, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Statistics in Medicine* 2005; 24: 1307-1319.

Turner 2000
Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* 2000; 19: 3417-3432.

Whitehead 2001
Whitehead A, Omar RZ, Higgins JPT, Savaluny E, Turner RM, Thompson SG. Meta-analysis of ordinal outcomes using individual patient data. *Statistics in Medicine* 2001; 20: 2243-2260.

ÍNDICE
CAPÍTULO 19: METANÁLISIS PROSPECTIVO

PUNTOS CLAVE

19.1 INTRODUCCIÓN

19.1.1 ¿Qué es un metanálisis prospectivo?
19.1.2 ¿Cuál es la diferencia entre un metanálisis prospectivo y un ensayo multicéntrico grande?
19.1.3 ¿Qué áreas de la atención sanitaria han utilizado el enfoque de metanálisis prospectivo?
19.1.4 ¿Qué recursos se necesitan?

19.2 LA NATURALEZA COLABORATIVA DE LOS METANÁLISIS PROSPECTIVOS

19.2.1 Grupos colaborativos
19.2.2 Negociación de la colaboración
19.2.3 Confidencialidad

19.3 EL PROTOCOLO DEL METANÁLISIS PROSPECTIVO

19.3.1 ¿Qué debe contener el protocolo?
Recuadro 19.3.a: Elementos de un protocolo de metanálisis prospectivo
19.3.2 Publicación del protocolo

19.4 OBTENCIÓN DE LOS DATOS EN EL METANÁLISIS PROSPECTIVO

19.5 ASPECTOS DEL ANÁLISIS EN EL METANÁLISIS PROSPECTIVO

19.5.1 Enfoque general
19.5.2 Análisis intermedio y monitorización de los datos

19.6 INFORMACIÓN DEL CAPÍTULO

Recuadro 19.6.a: El Grupo Cochrane de Métodos de Metanálisis Prospectivo

19.7 REFERENCIAS

CAPÍTULO 19: METANÁLISIS PROSPECTIVO

Autores: Davina Ghersi, Jesse Berlin y Lisa Askie del Grupo Cochrane de Métodos de Metanálisis Prospectivo.

Puntos clave

- Un metanálisis prospectivo es un metanálisis de estudios (habitualmente ensayos aleatorizados) que se identificaron, se evaluaron y se determinó que eran elegibles para el metanálisis antes de que se conocieran los resultados de cualquiera de ellos.
- Los metanálisis prospectivos permiten hacer hipótesis específicas por adelantado de los resultados de los ensayos individuales, permiten la aplicación prospectiva de los criterios de selección de los estudios y permiten declaraciones *a priori* de los análisis planificados. Los metanálisis, al contrario de los ensayos multicéntricos, admiten la variación de los protocolos de los estudios incluidos, lo que maximiza el poder estadístico en los metanálisis planificados previamente.
- Habitualmente los metanálisis prospectivos son realizados por un grupo colaborativo, que generalmente obtiene y analiza los datos de pacientes individuales.
- Los protocolos son importantes para los metanálisis prospectivos y se pueden publicar como protocolos para revisiones Cochrane. El Grupo Cochrane de Métodos de Metanálisis Prospectivo mantiene un registro de proyectos de metanálisis prospectivos y puede proporcionar asesoramiento para su realización.

19.1 Introducción

19.1.1 ¿Qué es un metanálisis prospectivo?

Una revisión sistemática realizada correctamente define la cuestión a abordar antes de la identificación de los ensayos potencialmente elegibles. Sin embargo, las revisiones sistemáticas son por naturaleza retrospectivas debido a que los ensayos incluidos habitualmente se identifican después que se han completado y han publicado sus resultados (Pogue 1998, Zanchetti 1998). El conocimiento de los resultados de los ensayos aleatorizados individuales puede introducir sesgo en una revisión sistemática retrospectiva si la selección de los componentes clave de la pregunta de revisión se basa en los artículos de uno o más ensayos positivos. Este hecho puede influir sobre:

- los criterios para la selección de los estudios (es decir, los tipos de ensayos considerados elegibles);
- la selección de la población objetivo;
- la naturaleza de la intervención;
- la selección del comparador; y
- los resultados a evaluar y sus medidas.

Tome, por ejemplo, una revisión sistemática en la cual los resultados de un estudio tienen una dirección opuesta a la de otros estudios en la revisión. Los revisores discuten las posibles explicaciones para esta heterogeneidad evidente y deciden que existe una explicación clínica. Sobre esta base, los revisores posteriormente deciden excluir el estudio. Ésta puede ser una decisión razonable; sin embargo, se hizo después de conocer el efecto de los resultados del estudio sobre la estimación resumen general y, por lo tanto, es intrínsicamente problemática.

Como se describe en detalle en el Capítulo 10 (Sección 10.2), el conocimiento de los resultados de un ensayo también puede influir en la decisión de publicar estos resultados. Incluso dentro de un ensayo publicado, los resultados pueden haber sido selectivamente descritos, lo que introduce una forma de sesgo de publicación más sutil en la revisión (Chan 2004).

Un metanálisis prospectivo (MAP) es un metanálisis de estudios (habitualmente ensayos aleatorizados) que se identificaron, se evaluaron y se determinó que eran elegibles para el metanálisis antes de que se conocieran los resultados de cualquiera de ellos. Los metanálisis prospectivos pueden tener características en común con los metanálisis acumulativos y con los que incluyen datos de pacientes individuales (Egger 1997). Los MAP pueden ayudar a superar algunos de los problemas reconocidos de los metanálisis retrospectivos (ver también Capítulo 18, Sección 18.5) ya que:

- permiten hacer hipótesis específicas por adelantado sin conocer los resultados de los ensayos individuales;
- permiten la aplicación prospectiva de los criterios de selección de los estudios; y
- permiten declaraciones a priori de los análisis planificados, incluidos los análisis de subgrupos, antes de que se conozcan los resultados de los ensayos individuales. Esto evita dificultades potenciales en la interpretación relacionada con el énfasis dependiente de los datos hecho en subgrupos específicos.

Las revisiones sistemáticas también dependen de la capacidad de los revisores para obtener los datos de todos los pacientes asignados al azar para los resultados relevantes, lo que puede ser difícil si no se proporcionó la información completa en las publicaciones del ensayo. Como la mayoría de los MAP obtendrán y analizarán datos de pacientes individuales (DPI) podrán superar este problema, con la ventaja adicional de poder

realizar análisis de tiempo hasta el evento de ser necesario. Los análisis de subgrupos planificados según los factores a nivel de paciente pueden proporcionar resultados erróneos si se cuenta solamente con los datos a nivel de conjunto, lo que destaca otra ventaja de los DPI. Los MAP también proporcionan una oportunidad única para que el diseño del ensayo, la obtención de los datos y otros procesos del ensayo clínico se estandaricen entre los ensayos. Por ejemplo, los investigadores pueden acordar utilizar el mismo instrumento para medir un resultado específico y medir el resultado en los mismos puntos temporales en cada ensayo. En una revisión Cochrane de intervenciones para prevenir la obesidad en niños, por ejemplo, la heterogeneidad y la falta de fiabilidad de algunos de los resultados medidos hace difícil combinar los datos entre los ensayos (Summerbell 2005). Un metanálisis prospectivo de esta cuestión ha propuesto un conjunto de normas generalmente comunes, de modo que se puedan abordar algunos de los aspectos surgidos debido a la falta de estandarización (Steinbeck 2006).

19.1.2 ¿Cuál es la diferencia entre un metanálisis prospectivo y un ensayo multicéntrico grande?

Los metanálisis prospectivos son una opción atractiva para los autores de ensayos clínicos que, aunque aprecian los beneficios de los ensayos únicos de tamaño adecuado, no pueden realizarlos (Simes 1987, Probstfield 1998). Ésta puede ser una metodología útil, por ejemplo, cuando se necesitan tamaños de muestra grandes para asegurar un poder estadístico adecuado, pero no es posible realizar ensayos únicos a gran escala. Esto se pudiera deber a intereses locales que impiden la participación en un ensayo cuando se percibe que la información se “pierde”. Este hecho también puede ser un problema específico en enfermedades poco frecuentes en las que puede ser difícil conseguir el acceso a un gran número de participantes en el ensayo de forma oportuna.

Por lo tanto, una alternativa para los investigadores es realizar localmente su propio estudio, colaborar con los investigadores de estudios similares y acordar que los resultados se combinen al completarse cada ensayo. Esto permite a los investigadores individuales mantener una cierta autonomía y al mismo tiempo planificar de forma apropiada el metanálisis. Otra situación en la que puede ser beneficioso, específicamente ante la falta de un registro prospectivo obligatorio de ensayos aleatorizados, es cuando se comienzan dos o más ensayos que abordan la misma cuestión y los investigadores ignoran la existencia del/los otro/s ensayo/s. Una vez que se identifican ensayos similares, los investigadores pueden colaborar (y adaptar la obtención de datos, si es necesario) y planificar prospectivamente la combinación de sus resultados en un metanálisis.

Algo que también distingue un MAP de un ensayo multicéntrico es que en un MAP no se necesita que los protocolos sean idénticos entre los estudios. La variedad en el diseño de los estudios puede ser observada por algunos como una característica conveniente de los MAP, por lo que se considera aceptable un grado de variación natural en las poblaciones o en aspectos de las intervenciones. FICSIT (*Frailty and Injuries: Cooperative Studies of Intervention Techniques*) es un ejemplo de un metanálisis planificado previamente de ocho estudios de intervenciones con ejercicios en una población débil de edad avanzada (Schechtman 2001). Los ocho sitios FICSIT definieron sus propias intervenciones mediante criterios específicos adaptados a la situación local y evaluaciones y difirieron en los criterios de ingreso (excepto en que todos los participantes fueron personas de edad avanzada). Esta introducción deliberada de variabilidad sistemática en el diseño, conocida como “diseño metaexperimental” es un posible enfoque para los MAP (Cholesterol Treatment Trialists’ [CTT] Collaboratos 2005).

19.1.3 ¿Qué áreas de la atención sanitaria han utilizado el enfoque de metanálisis prospectivo?

El metanálisis prospectivo es un método que han utilizado en años recientes los autores de ensayos en enfermedades cardiovasculares (Simes 1995, WHO - ISI Blood Pressure Lowering Treatment Trialist' Collaboration 1998), leucemia infantil (Shuster 1996, Valsecchi 1996) y obesidad en la infancia y la adolescencia (Steinbeck 2006). Además, se han identificado algunas áreas como las enfermedades infecciosas en las que en gran medida se ha pasado por alto la oportunidad de utilizar un MAP (Ioannidis 1999). El sitio web del Grupo Cochrane de Métodos de Metanálisis Prospectivo incluye una lista de MAP en curso o completados en los que se puede encontrar información adicional (Ghersi 2005).

19.1.4 ¿Qué recursos se necesitan?

Los MAP son tareas considerables y no se deberían emprender a la ligera. Es probable que un MAP lleve muchos años completar y necesita un secretariado adecuadamente financiado y con un personal apropiado, comprometido y activo. Una vez que se ha conformado el grupo colaborativo (ver Sección 19.2), se necesitan recursos para asegurar el compromiso activo del grupo durante muchos años, con frecuencia un período de tiempo muchos más largo que el que se necesita para una revisión retrospectiva de DPI (ver Capítulo 18). El secretariado necesitará organizar teleconferencias regulares, reuniones presenciales (al menos de forma anual), informes, actualizar los detalles de contacto e implementar otros mecanismos para mantener unido al grupo colaborativo. Este tipo de actividad es semejante a la que realiza el centro coordinador de un ensayo aleatorizado multicéntrico. Uno de los beneficios de estas actividades del secretariado es que con frecuencia ayudan a facilitar el cumplimiento con el protocolo del MAP y estimulan el seguimiento completo en los ensayos individuales participantes.

19.2 La naturaleza colaborativa de los metanálisis prospectivos

19.2.1 Grupos colaborativos

Como sucede con los metanálisis de DPI (ver Capítulo 18, Sección 18.2.1), la mayoría de los MAP los realizan y publican los grupos colaborativos. Los grupos colaborativos deberían incluir representantes de cada uno de los ensayos participantes y habitualmente tendrán un grupo directivo o secretariado que dirige el proyecto diariamente. El grupo colaborativo puede optar por crear grupos *ad hoc* pequeños para tratar aspectos específicos que surjan y proporcionar asesoramiento al grupo directivo o al secretariado sobre aspectos clínicos, técnicos o de otro tipo que puedan repercutir en el proyecto.

19.2.2 Negociación de la colaboración

Como ocurre con los metanálisis de DPI (ver Capítulo 18, Sección 18.2.2), la negociación y el establecimiento de una colaboración sólida con los autores de los ensayos participantes es esencial para el éxito de un MAP. Sin embargo, el objetivo de un MAP no es principalmente la localización y obtención de los datos de los ensayos individuales. Como es necesario formar la colaboración antes de que se conozcan los resultados de cualquier ensayo, el objetivo de los esfuerzos colaborativos de un MAP, al menos al inicio, es alcanzar un acuerdo con respecto a la población de estudio, el diseño y los métodos de obtención de los datos para cada uno de los estudios participantes. Cuando

los miembros de un grupo colaborativo de un MAP consienten participar en el proyecto, necesitan acordar un protocolo central común e ítems de datos centrales comunes que se recopilarán entre los ensayos. Los ensayos individuales pueden incluir enmiendas al protocolo local o a ítems de datos adicionales, pero deberían asegurar que éstos no comprometerán los elementos del protocolo central común.

En un MAP se realizan esfuerzos para identificar *todos* los ensayos en curso para maximizar la precisión y evitar los sesgos que se pudieran introducir por la exclusión de estudios (al menos en parte) debido al conocimiento de los resultados de los mismos. Para certificar que un estudio individual es elegible para inclusión en el MAP deberían existir evidencia que apoye la demanda de que, en el momento del acuerdo de formar parte del MAP, los resultados del ensayo no se conocían fuera del comité de monitorización de los datos del propio ensayo. Idealmente ésta debería ser la forma de probar que el ensayo se registró de forma prospectiva (Laine 2007). También sería conveniente para el grupo colaborativo obtener un acuerdo explícito (y firmado) de cada uno de los grupos de los ensayos que van a colaborar. La idea es estimular contribuciones significativas de los investigadores individuales y lograr que se “apropien” del concepto del MAP y de los detalles del protocolo.

19.2.3 Confidencialidad

Los aspectos de la confidencialidad con respecto al anonimato y la seguridad de los datos son similares a los descritos en el Capitulo 18 (Sección 18.2.3) para los metanálisis de DPI. Los aspectos específicos para los MAP incluyen la planificación adecuada con respecto a cómo tratar los ensayos incluidos en el MAP que finalicen y publiquen sus resultados, y cómo tratar los aspectos relacionados con la monitorización y la seguridad de los datos, incluidos la repercusión de los análisis intermedios de los ensayos individuales en el MAP, o posiblemente un análisis intermedio combinado del MAP (ver también Sección 19.5.2).

19.3 El protocolo del metanálisis prospectivo

19.3.1 ¿Qué debe contener el protocolo?

Todos los MAP deberían tener un protocolo disponible públicamente. El desarrollo de un protocolo para un MAP es similar, conceptualmente, al realizado para un ensayo único. Los elementos esenciales para un MAP se detallan a continuación y se resumen en el Recuadro 19.3.a.

Objetivos, elegibilidad y resultados

Como en cualquier protocolo, el primer paso importante es definir la hipótesis y luego establecer los criterios de elegibilidad para los estudios. Por ejemplo, pudiera ser necesario que los estudios incluidos en el MAP utilizaran una asignación aleatoria de los participantes a las intervenciones, aunque es posible incluir otros diseños de estudios en un MAP. Los ensayos individuales, si son aleatorizados, pueden elegir compartir un método de asignación al azar común, o al menos utilizar los mismos factores de estratificación. Es necesario que se especifiquen los atributos de la población participante, así como el mínimo de requisitos para cada una de las intervenciones y los brazos comparadores. El protocolo también debiera especificar qué resultados es necesario medir, cuándo y cómo se deberían medir, cuáles son primarios y cuáles secundarios, así como otras características del diseño del estudio cuando sea necesario. Si un MAP se establece desde el inicio, puede ser posible para cada ensayo en el MAP compartir exactamente el mismo protocolo de ensayo.

Métodos de búsqueda

El protocolo debería describir en detalle los esfuerzos realizados para identificar los ensayos en curso e incluir cómo han sido (o serán) localizados y abordados los colaboradores potenciales para solicitar su participación.

Detalles del ensayo

En el protocolo se deberían enumerar los detalles de los ensayos identificados para inclusión (si son relevantes). El listado pudiera incluir el número anticipado de participantes y el plazo definido para cada ensayo participante. El protocolo debería incluir una declaración general de si, en el momento de la presentación para el registro, se conocía cualquiera de los resultados del ensayo (por cualquier persona fuera del comité de monitorización de los datos del propio ensayo). Los ensayos se deberían incluir sólo si no se conocen sus resultados en el momento en el cual los mismos se identifican e incluyen en el MAP. Si los ensayos elegibles se identificaron pero no se incluyeron en el MAP porque sus resultados ya se conocían, el protocolo del MAP debería dar una idea de qué se hará con ellos. Por ejemplo, pudieran realizarse análisis de sensibilidad secundarios que utilicen los datos de pacientes individuales o grupales de estos ensayos. El protocolo debería describir las acciones a realizar si posteriormente se localizan ensayos mientras el MAP está en progreso.

Plan de análisis

El protocolo debería esbozar los planes para la obtención y los análisis de los datos de forma similar a la de los metanálisis de DPI (ver Capítulo 18). Éste incluiría detalles del tamaño de la muestra y el cálculo del poder estadístico (para el MAP), cualquier análisis intermedio a realizar y detalles de los análisis de subgrupos planificados. También se pueden incorporar en un MAP las estrategias para abordar cuestiones adicionales más allá de la hipótesis principal de interés. Estas cuestiones adicionales se pueden agregar mientras los resultados de los estudios que se incluirán en el análisis no se conozcan, es decir, no sean preguntas de investigación “condicionadas por los datos”. Es de señalar que puede haber análisis que sean únicos para el MAP que *no* se realizan dentro de los ensayos individuales, como los análisis de subgrupos.

A los investigadores de los ensayos que se incluirán en un MAP generalmente se les debería consultar su conformidad para proporcionar los datos de pacientes individuales. El protocolo debería describir qué ocurriría si los investigadores de alguno de los estudios incluidos en el MAP no pueden (o no desean) proporcionar los datos a nivel de paciente, quizás debido a aspectos relacionados con la confidencialidad o el consentimiento informado. ¿Aceptaría el secretariado del MAP, por ejemplo, datos resumidos adecuados? (Se pudiera realizar un análisis de dos etapas, en el cual la estimación del efecto de interés se calcule de forma separada dentro de cada estudio y se utilicen los datos a nivel de paciente, y entonces las estimaciones dentro del estudio se combinen entre los estudios mediante métodos metanalíticos estándar). El protocolo debería especificar si se pretende actualizar los datos del MAP a intervalos regulares mediante ciclos continuos de obtención de datos (p.ej. cada cinco años) y por lo tanto, cuando es de esperar que los autores de los ensayos suministren datos de resultados a lago plazo actualizados.

Procesamiento y coordinación

El protocolo del MAP debería esbozar los detalles de la estructura de la dirección del proyecto (incluido cualquier comité, ver Sección 19.2.1), los procedimientos para el procesamiento de los datos (cómo se obtendrán los datos, el formato requerido, cuándo se requerirá presentar los datos, calidad de los procedimientos de garantía, etc.; ver Capítulo 18, Sección 18.3) y quiénes serán responsables de los análisis estadísticos.

Política de publicación

Un elemento clave del protocolo del MAP es la política de publicación. Es fundamental tener una política con respecto a la autoría (p.ej. especificar que la publicación será a

nombre del grupo, pero también incluye una lista de autores individuales). También es importante una política con respecto a la preparación del manuscrito. Por ejemplo, se pudiera especificar que los borradores de los artículos circulen entre todos los autores de los ensayos antes de la presentación para la publicación. Pudiera haber un comité de redacción, como los que frecuentemente se forman dentro de los grupos de estudios cooperativos.

Un aspecto específico que surge en el contexto del MAP (que generalmente no aparece en un estudio multicéntrico ni en un metanálisis de DPI) es si los estudios individuales deberían o no publicar sus propios datos y el momento de dicha publicación. La mayoría de los investigadores desearían publicar sus propios estudios de forma individual además de contribuir al MAP, y es probable que los investigadores deseen que estas publicaciones aparezcan antes que el MAP sea publicado, lo que evita los aspectos relacionados con la publicación duplicada de los mismos datos. No obstante, con el mismo objetivo, cualquier publicación del MAP debería indicar claramente las fuentes de los datos incluidos y nombrar las publicaciones anteriores de los mismos datos. El protocolo del MAP también debería señalar qué ocurriría si cualquiera de los ensayos participantes no publica sus resultados individuales dentro del marco de tiempo especificado. Esto pudiera ocurrir si el ensayo no es completado debido a financiamiento insuficiente, se termina de forma prematura o simplemente permanece sin publicarse después de la fecha preespecificada. El protocolo también debería abordar cómo se procederá con los ensayos que no cumplan con su acuerdo de participar en el MAP.

Recuadro 19.3.a: Elementos de un protocolo de metanálisis prospectivo

Objetivos:
- Definir las hipótesis/objetivos específicos.

Métodos: Criterios para considerar los estudios para esta revisión:
- Criterios de elegibilidad para el diseño del ensayo (p.ej. requisitos para la asignación al azar, seguimiento mínimo).
- Criterios de elegibilidad para la población de pacientes.
- Criterios de elegibilidad para cada intervención y comparador.
- Información de los resultados: especificación de las variables de evaluación primarias y secundarias, definiciones, instrumentos de medición, momento.
- Detalles de los subgrupos.

Métodos: Métodos de búsqueda para la identificación de los estudios:
- Describir los esfuerzos para identificar los ensayos en curso.

Métodos: Obtención y análisis de los datos:
- Detalles del ensayo:
  - Lista de detalles de los ensayos identificados para inclusión.
  - Una declaración general de si, en el momento de la presentación del MAP para el registro, se conocía cualquiera de los resultados del ensayo (por cualquier persona fuera del comité de monitorización de los datos del propio ensayo). Los ensayos se deberían incluir solamente si en el momento en que se identificaron y agregaron al MAP se desconocían sus resultados.
  - Si se obtuvo un acuerdo firmado para colaborar de los representantes apropiados de cada ensayo (p.ej. Patrocinador o Investigador principal).
- Plan de análisis:
  - Detalles del tamaño de la muestra y el cálculo del poder estadístico (para los MAP), análisis intermedios, análisis de subgrupos, etc.
- Procesamiento y coordinación:
  - Detalles de la estructura y los comités de procesamiento.
  - Procesamiento de los datos (datos a obtener, formato requerido, cuando se requirió, procedimientos para asegurar la calidad, etc.).
  - Responsabilidad de los análisis estadísticos.
- Política de publicación:
  - Política con respecto a la autoría (publicación a nombre de un “grupo”).
  - Comité de Redacción (afiliación, responsabilidades).
  - Política con respecto al texto (p.ej. se circuló a todos los autores de los ensayos para que realicen comentarios).

19.3.2 Publicación del protocolo

Si se preparó como una revisión Cochrane, el protocolo del MAP se debería presentar al Grupo Cochrane de Revisión apropiado para que aparezca en la *Base de Datos Cochrane de Revisiones Sistemáticas*. Por otra parte, un protocolo puede ser publicado en cualquier otra parte (por ejemplo, el Protocolo CTT/PPP [Cholesterol Treatment Trialist' (CTT) Collaborators 2005]). También es conveniente que los proyectos MAP se registren en el sitio web del Grupo Cochrane de Métodos de Metanálisis Prospectivos (ver Recuadro 19.6.a) y la información acerca del proyecto se debería actualizar al menos anualmente. Cada ensayo incluido en el MAP se debería registrar en un Registro Primario, públicamente accesible, reconocido por la OMS (www.who.int/ictrp/network/list_registers) antes del reclutamiento del primer participante, y de acuerdo con los requisitos internacionales (Sim 2006, Laine 2007).

19.4 Obtención de los datos en el metanálisis prospectivo

Habitualmente los autores de los ensayos que participan en un MAP suministran los datos de pacientes individuales una vez que sus ensayos se han completado y publicado. La ventaja del diseño del MAP es que los autores de los ensayos deciden de forma prospectiva qué datos se obtendrán y en qué formato, por lo cual la necesidad de redefinir y recodificar los datos suministrados es menos problemática, lo que frecuentemente ocurre con los DPI retrospectivos. El MAP debería desarrollar un protocolo de transferencia de datos que pueda incorporar normas actuales de intercambio de datos como las desarrolladas por el *Clinical Data Interchange Standard Consortium* (CDISC; www.cdisc.org).

Una vez que el secretariado del MAP recibe los datos, debería verificarlos de forma cuidadosa mediante los mismos procedimientos utilizados en los metanálisis de DPI, lo que incluye la verificación de los datos faltantes o duplicados, la verificación de la verosimilitud de los datos, la evaluación de los patrones de asignación al azar y la confirmación de que la información suministrada está actualizada (ver Capítulo 18, Sección 18.4.4). Las dudas sobre los datos se deberían obtener mediante consulta directa con los autores de los ensayos individuales antes de incluirlos en el conjunto final de datos para el análisis.

19.5 Aspectos del análisis en el metanálisis prospectivo

19.5.1 Enfoque general

La mayoría de los MAP utilizan técnicas de análisis general similares a la de los metanálisis retrospectivos de DPI. Estas técnicas se describen detalladamente en el Capítulo 18 (Sección 18.4) e incluyen el enfoque general de estos análisis y la posibilidad de realizar análisis de tiempo hasta el evento (si son apropiados). El uso de los datos a nivel del paciente también permite análisis de subgrupos estadísticamente más poderosos y un modelado multinivel para explorar las asociaciones entre los efectos de la intervención y las características de los pacientes, así como un modelado pronóstico en algunos casos. El Capítulo 18 (Sección 18.4.7) describe algunos de los paquetes de programas informáticos que posiblemente se puedan utilizar para analizar estos tipos de datos.

19.5.2 Análisis intermedio y monitorización de los datos

En la práctica cada vez es más frecuente que los ensayos clínicos individuales incluyan un plan de análisis intermedio de los datos y una monitorización de la seguridad. Los MAP ofrecen una oportunidad única para realizar estos vistazos intermedios utilizando los datos aportados por todos los ensayos. Para este análisis los datos se pueden combinar o se pueden estudiar por separado para cada ensayo, y luego compartir los resultados entre los comités de monitorización de datos de los ensayos participantes.

La capacidad para realizar análisis intermedios da lugar a varios problemas éticos. Por ejemplo, ¿es apropiado continuar la asignación al azar en los estudios en curso después que se ha demostrado un beneficio general (en términos de la variable principal, por ejemplo) de una intervención? Cuando se desconocen los resultados en los subgrupos de interés clínico o para las variables menos frecuentes, los investigadores deberían proseguir con el estudio hasta obtener información adicional sobre el beneficio clínico neto general, por ejemplo, evidencia de beneficio para un resultado pero no para otro o evidencia de daño.

Si cada ensayo tiene su propio comité de monitorización de datos, entonces la comunicación entre los comités pudiera ser beneficiosa en este aspecto, como recomiendan Hillman y Louis (Hillman 2003). Los diferentes comités necesitarían estar al tanto de los otros ensayos incluidos dentro del MAP y de sus resultados, ya que estas consideraciones externas pudieran influir en las decisiones tomadas por un comité de monitorización; por ejemplo, si un estudio se detiene temprano debido a evidencia de eficacia. Por el contrario, se pudiera indicar que el conocimiento de los datos de seguridad emergentes de todos los ensayos participantes pudiera reducir las posibilidades de interrupciones tempranas falsas de un ensayo individual debido a aspectos relacionados con resultados de seguridad intermedios. Pudiera ser de ayuda, por lo tanto, que los comités de seguridad de los datos de ensayos diferentes adopten un acuerdo común de que los ensayos individuales no se deberían interrumpir hasta que no se logren los objetivos del MAP relacionados con los subgrupos y las variables de evaluación poco frecuentes (o “beneficio clínico neto”).

Otra opción posible pudiera ser considerar limitar el reclutamiento, en los ensayos que continúan, a los pacientes en el/los subgrupo/s de interés, si tal decisión tiene un sentido clínico y estadístico. En cualquier caso, pudiera ser apropiado aplicar los conceptos de la metodología de ensayos clínicos secuenciales, como el enfoque descrito por Whitehead (Whitehead 1997), para obtener reglas de detención rigurosas y severas para el MAP cuando los resultados de los ensayos individuales estén disponibles.

19.6 Información del capítulo

Autores: Davina Ghersi, Jesse Berlin y Lisa Askie del Grupo Cochrane de Métodos de Metanálisis Prospectivo.

La versión en inglés de este capítulo se debería citar como: Ghersi D, Berlin J, Askie L. Chapter 19: Prospective meta-analysis. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Recuadro 19.6.a: El Grupo Cochrane de Métodos de Metanálisis Prospectivo

La función del Grupo de Métodos de Metanálisis Prospectivo (GMMAP) es:

- Proporcionar un mecanismo que permita el registro de los metanálisis prospectivos:
  - Cochrane (mediante los Grupos de Revisión); y
  - no Cochrane (mediante GMMAP);
- Proporcionar un mecanismo para evaluar los protocolos enviados para su registro para asegurar que los mismos sean en efecto metanálisis prospectivos. Esto se puede lograr de la forma siguiente:
  - proporcionar formación a los miembros de los Grupos Cochrane de Revisión (p.ej. editores y revisores externos);
  - que los miembros del GMMAP sean revisores externos de los protocolos; y
  - mediante una lista de verificación para los investigadores que realizan o son revisores externos de un MAP;
- Desarrollar normas metodológicas apropiadas para los metanálisis prospectivos.
- Proporcionar asesoramiento y apoyo a quienes emprenden (o tienen la intención de emprender) metanálisis prospectivos.

Los miembros del grupo admiten a cualquiera que realice, haya realizado o esté interesado en realizar un metanálisis prospectivo, independientemente del área de atención sanitaria investigada. Para afiliarse, se pide a los individuos detallar su nivel de compromiso en un cuestionario del Grupo de Métodos de Metanálisis Prospectivo (disponible en el sitio web de MAP, ver más adelante). A los miembros se les solicitará actualizar esta información anualmente.

Sitio web: www.cochrane.org/docs/pma.htm

19.7 Referencias

Chan 2004
Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004; 291: 2457-2465.

Cholesterol Treatment Trialists' (CTT) Collaborators 2005
Cholesterol Treatment Trialists' (CTT) Collaborators. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90 056 participants in 14 randomised trials of statins. *The Lancet* 2005; 366: 1267-1278.

Egger 1997
Egger M, Davey Smith G. Meta-analysis: potentials and promise. *BMJ* 1997; 315: 1371-1374.

Ghersi 2005
Ghersi D. Cochrane Prospective Meta-analysis Methods Group. *About the Cochrane Collaboration (Methods Groups)* 2005, Issue 2. Art No: CE000132.

Hillman 2003
Hillman DW, Louis TA. DSMB case study: decision making when a similar clinical trial is stopped early. *Controlled Clinical Trials* 2003; 24: 85-91.

Ioannidis 1999
Ioannidis JPA, Lau J. State of the evidence: current status and prospects of meta-analysis in infectious diseases. *Clinical Infectious Diseases* 1999; 29: 1178-1185.

Laine 2007
Laine C, Horton R, DeAngelis CD, Drazen JM, Frizelle FA, Godlee F, Haug C, Hebert PC, Kotzin S, Marusic A, Sahni P, Schroeder TV, Sox HC, Van der Weyden MB, Verheugt FW. Clinical trial registration: looking back and moving ahead. *Canadian Medical Association Journal* 2007; 177: 57-58.

Pogue 1998
Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomissed controlled trials. *The Lancet* 1998; 351: 47-52.

Probstfield 1998
Probstfield J, Applegate WB. Prospective meta-analysis: Ahoy! A clinical trial? *Journal of the American Geriatrics Society* 1988; 43: 452-453.

Schechtman 2001
Schechtman K, Ory M. The effects of exercise on the quality of life of frail older adults: a preplanned meta-analysis of the FICSIT trials. *Annals of Behavioural Medicine* 2001; 23: 186-197.

Shuster 1996
Shuster JJ, Gieser PW. Meta-analysis and prospective meta-analysis in childhood leukemia clinical research. *Annals of Oncology* 1996; 7: 1009-1014.

Sim 2006
Sim I, Chan AW, Gulmezoglu M, Evans T, Pang T. Clinical trial registration: transparency is the watchword. *The Lancet* 2006; 367: 1631-1633.

Simes 1987
Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Statistics in Medicine* 1987; 6: 11-29.

Simes 1995
Simes RJ. Prospective meta-analysis of cholesterol-lowering studies: the Prospective Pravastatin Pooling (PPP) Project and the Cholesterol Treatment Trialists' (CTT) Collaboration. *American Journal of Cardiology* 1995; 76: 122c-126c.

Steinbeck 2006
Steinbeck KS, Baur LA, Morris AM, Ghersi D. A proposed protocol for the development of a register of trials of weight management of childhood overweight and obesity. *International Journal of Obesity* 2006; 30: 2-5.

Summerbell 2005
Summerbell CD, Waters E, Edmunds LD, Kelly S, Brown T, Campbell KJ. Interventions for preventing obesity in children. *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art No: CD001871.

Valsecchi 1996
Valsecchi MG, Masera G. A new challenge in clinical research in childhood ALL: the prospective meta-analysis strategy for intergroup collaboration. *Annals of Oncology* 1996; 7: 1005-1008.

Whitehead 1997
Whitehead A. A prospectively planned cumulative meta-analysis applied to a series of concurrent clinical trials. *Statistics in Medicine* 1997; 16: 2901-2913.

WHO - ISI Blood Pressure Lowering Treatment Trialists' Collaboration 1998
WHO - ISI Blood Pressure Lowering Treatment Trialists' Collaboration. Protocol for prospective collaborative overviews of major randomised trials of blood-pressure-lowering treatments. *Journal of Hypertension* 1998; 16: 127-137.

Zanchetti 1998
Zanchetti A, Mancia G. Searching for information from unreported trials - amnesty for the past and prospective meta-analysis for the future. *Journal of Hypertension* 1998; 16: 125.

ÍNDICE

CAPÍTULO 20: INVESTIGACIÓN CUALITATIVA Y REVISIONES COCHRANE

PUNTOS CLAVE

20.1 INTRODUCCIÓN

20.2 INCORPORACIÓN DE LA EVIDENCIA DE LA INVESTIGACIÓN CUALITATIVA EN LAS REVISIONES COCHRANE DE INTERVENCIONES: CONCEPTOS Y ASPECTOS

20.2.1 Definición de investigación cualitativa
20.2.2 Uso de la evidencia proveniente de la investigación cualitativa en las revisiones Cochrane
20.2.3 Consideración de los estudios cualitativos que se identifican en los ensayos controlados aleatorizados o junto a ellos
20.2.4 Consideraciones relacionadas con los recursos

20.3 SÍNTESIS DE LA EVIDENCIA CUALITATIVA

20.3.1 Ejemplo de síntesis de evidencia cualitativa para complementar una revisión Cochrane de intervenciones: tratamiento observado directamente y tuberculosis (TB)
Recuadro 20.3.a: Tratamiento directamente observado de la tuberculosis: una síntesis del resumen de la evidencia cualitativa
20.3.2 Aspectos metodológicos
20.3.2.1 Estrategias de búsqueda
20.3.2.2 Evaluación crítica
20.3.2.3 Síntesis de la evidencia de la investigación cualitativa
20.3.2.4 Selección del método apropiado
20.3.2.5 Enfoques para integrar la síntesis de la evidencia cualitativa y cuantitativa
20.3.2.6 Conclusión

20.4 INFORMACIÓN DEL CAPÍTULO

Recuadro 20.4.a: El Grupo Cochrane de Métodos de Investigación Cualitativa

20.5 REFERENCIAS

20.6 FURTHER SELECTED READING

20.6.1 Qualitative research, general
20.6.2 Qualitative methods
20.6.3 Qualitative literature searching
20.6.4 Synthesizing qualitative evidence
20.6.5 Synthesizing qualitative and quantitative evidence
20.6.6 Critical appraisal of qualitative studies
20.6.7 Web sites

CAPÍTULO 20: INVESTIGACIÓN CUALITATIVA Y REVISIONES COCHRANE

Autores: Jane Noyes, Jennie Popay, Alan Pearson, Karin Hannes y Andrew Booth del Grupo Cochrane de Métodos de Investigación Cualitativa.

Puntos clave

- La evidencia proveniente de los estudios cualitativos puede tener una función importante para agregarle valor a las revisiones sistemáticas sobre toma de decisiones políticas, prácticas y por parte de los consumidores.
- Es probable que los estudios de resultados incluidos en las revisiones Cochrane tengan incorporadas investigaciones cualitativas o se asocien con ellas.
- La investigación cualitativa puede contribuir con las revisiones Cochrane de intervenciones de cuatro formas:

    a. informar a las revisiones a través del uso de investigaciones cualitativas para ayudar a definir y refinar la pregunta, y asegurar que la revisión incluya estudios apropiados y aborde resultados importantes;
    b. mejorar las revisiones al sintetizar la evidencia de la investigación cualitativa identificada, a la vez que buscan evidencia de efectividad;
    c. ampliar las revisiones al realizar una búsqueda para investigar específicamente la evidencia de los estudios cualitativos que aborden preguntas directamente relacionadas con la revisión de efectividad; y
    d. complementar las revisiones al sintetizar las pruebas cualitativas dentro de una revisión cualitativa independiente, o complementaria, que aborde preguntas sobre aspectos que no sean la efectividad.
- Existen muchos métodos de síntesis de la evidencia cualitativa apropiados a los objetivos y alcance de las revisiones Cochrane de intervenciones.
- La síntesis de la investigación cualitativa es una área de debate y en evolución. El Grupo Cochrane de Métodos Cualitativos proporciona un foro para la discusión y el desarrollo adicional de la metodología en esta área.

20.1 Introducción

El objetivo de este capítulo es esbozar las formas en las cuales se puede utilizar la investigación cualitativa para informar, mejorar, ampliar y complementar las revisiones Cochrane. Las pruebas cualitativas no pretenden contribuir a las medidas del efecto de las intervenciones, sino en su lugar ayudar a explicar, interpretar y aplicar los resultados de una revisión Cochrane. De esta manera, la evidencia derivada de los estudios cualitativos complementa una revisión sistemática de estudios cuantitativos.

Este capítulo tiene como objetivo permitir a los revisores:

1. considerar los tipos de revisiones y preguntas de revisión para los cuales una síntesis de las pruebas cualitativas pudiera mejorar o ampliar una revisión Cochrane;
2. considerar los aspectos de recursos y metodológicos cuando se decide sintetizar evidencia cualitativa para complementar una revisión Cochrane;
3. llamar la atención sobre algunos de los enfoques y métodos disponibles para la síntesis de pruebas cualitativas; y
4. acceder a información, asesoramiento y recursos adicionales, de ser necesario.

El capítulo se divide en dos partes. La primera parte (Sección 20.2) proporciona algunas consideraciones y pautas para la incorporación de evidencia proveniente de investigaciones cualitativas en las revisiones Cochrane, incluidas las repercusiones de recursos. La segunda parte (Sección 20.3) proporciona una discusión más general de los aspectos metodológicos, las lecturas clave y la función y los detalles del Grupo Cochrane de Métodos de Investigación Cualitativa. Se proporciona un ejemplo que muestra cómo se ha utilizado una síntesis de la evidencia cualitativa para complementar una revisión Cochrane de efectos ya existente.

20.2 Incorporación de la evidencia de la investigación cualitativa en las revisiones Cochrane de intervenciones: conceptos y aspectos

20.2.1 Definición de investigación cualitativa

Los investigadores cualitativos estudian eventos en su ámbito natural e intentan hallar el sentido o interpretar los fenómenos en términos de los significados que las personas les dan (Denzin 1994). La investigación cualitativa pretende penetrar en la significación más profunda que el sujeto de investigación otorga al tema que se investiga. Esto incluye un enfoque interpretativo y naturalístico del tema y le da prioridad a la contribución de los datos a importantes preguntas de investigación o información existente.

Dentro de la atención sanitaria, la comprensión del valor de la evidencia de la investigación cualitativa para las revisiones sistemáticas debería considerar la naturaleza variada y difusa de la evidencia (Popay 1998b, Pearson 2005). La investigación cualitativa incluye un rango de filosofías, diseños de investigación y técnicas específicas que incluyen las entrevistas en profundidad, la observación de participantes y no participantes; los grupos focales; los análisis de documentos y otros múltiples métodos de obtención de datos (Pope 2006). Debido a esta variedad de tipos de datos, también existen diferentes enfoques teóricos y metodológicos para los diseños de estudios y el análisis de los datos como la

fenomenología, la etnografía, la teoría de base (*grounded theory*), la investigación acción, los estudios de casos y muchos otros. La teoría y las perspectivas de los investigadores también tienen una función clave en el análisis de los datos cualitativos y en las bases sobre las cuales se pueden hacer generalizaciones a otros contextos.

Dentro de las ciencias empíricas, la posición de una teoría determinada depende completamente de la cantidad y el carácter de la evidencia a su favor. Es la ponderación relativa de la evidencia en la que se apoya lo que permite seleccionar entre teorías opuestas. Dentro de las ciencias naturales la generación del conocimiento incluye probar una hipótesis o un grupo de hipótesis al derivar consecuencias de ellas y luego probar si estas consecuencias se mantienen ciertas después del experimento y la observación.

Los profesionales de la salud buscan pruebas que apoyen el valor de un amplio rango de actividades e intervenciones, por lo que el tipo de pruebas que se necesitan dependen de la naturaleza de la actividad y su objetivo. Para muchas preguntas, por ejemplo, las relacionadas con las creencias de los padres y los niños acerca de la vacunación (Mills 2005a, Mills 2005b), la investigación cualitativa es una metodología apropiada y conveniente.

20.2.2 Uso de la evidencia proveniente de la investigación cualitativa en las revisiones Cochrane

Las revisiones Cochrane de intervenciones se dirigen principalmente a determinar si una intervención es efectiva comparada con un control y, de serlo, calcular el tamaño del efecto. Los ensayos aleatorizados de alta calidad son fundamentales para los objetivos de la Colaboración Cochrane con respecto a este tema. No es apropiado ni posible incluir evidencia proveniente de la investigación cualitativa en todas las revisiones Cochrane.

Sin embargo, cada vez se reconoce más que la evidencia de los estudios cualitativos que explora la experiencia de las personas involucradas en proporcionar y recibir intervenciones, y los estudios que evalúan los factores que influyen en la implementación de las intervenciones tienen una función importante en asegurar que las revisiones sistemáticas tengan un valor máximo para la toma de decisiones políticas, prácticas y por parte de los consumidores (Mays 2005, Arai 2005, Popay 2005).

La relevancia de la evidencia cualitativa para la evaluación de las intervenciones apenas recientemente ha recibido el reconocimiento en el campo de la salud, pero actualmente es más frecuente que los componentes cualitativos se incluyan en la evaluación de las intervenciones sanitarias (Pope 2006) y que para la evaluación de intervenciones complejas como los diferentes modelos de administración se utilicen enfoques de “métodos mixtos”. Por lo tanto, cada vez es más probable que los estudios de resultados incluidos en las revisiones Cochrane tengan incorporadas investigaciones cualitativas o se asocien con ellas. Por este motivo, los autores de revisiones Cochrane preguntan con mayor frecuencia cómo utilizar la evidencia de la investigación cualitativa para mejorar la relevancia y la utilidad de sus revisiones para posibles usuarios.

Una síntesis de la evidencia proveniente de la investigación cualitativa puede explorar preguntas como ¿en qué forma las personas desarrollan la enfermedad?, ¿por qué una intervención funciona (o no)?, ¿para quién y en qué circunstancias? En algunas revisiones, particularmente las que abordan la provisión de atención sanitaria, puede ser conveniente contar con pruebas cualitativas que aborden preguntas como ¿cuáles son las barreras y los facilitadores para acceder a la atención sanitaria? O ¿cuál es la repercusión que las barreras y los facilitadores específicos tienen sobre las personas, sus experiencias y sus conductas? Estas se pueden generar, por ejemplo, a través de estudios etnográficos y de entrevistas de conductas de búsqueda de ayuda. Las pruebas provenientes de la investigación cualitativa

pueden ayudar a la interpretación de los resultados de las revisiones sistemáticas al ayudar a comprender la forma en la cual la intervención es percibida por los que participan en el desarrollo, la provisión y la recepción de la misma; qué aspectos de la intervención ellos valoran o no; y por qué es así. Estos tipos de evidencia cualitativa puede proporcionar detalles de los factores que son externos a una intervención, incluidos, por ejemplo, la repercusión de otros avances en la política, los factores que facilitan o entorpecen la ejecución exitosa de un programa, servicio o tratamiento y cómo una intervención específica puede necesitar adaptarse para una implementación a gran escala (Roen 2006).

Se han identificado cuatro formas en las cuales la investigación cualitativa puede contribuir con las revisiones Cochrane para la política y la práctica sanitarias (Popay 2006a):

1. Informar a las revisiones a través del uso de investigación cualitativa para ayudar a definir y refinar la pregunta. Esto asegura que la revisión incluya estudios apropiados y aborde resultados importantes, lo que permite que la revisión sea de máxima relevancia para los posibles usuarios.
2. Mejorar las revisiones al sintetizar la evidencia de la investigación cualitativa identificada, a la vez que buscan pruebas de efectividad. La evidencia cualitativa asociada con los ensayos se puede utilizar para explorar aspectos de la implementación de la intervención. En la Sección 20.2.3 se analiza en más detalle la investigación cualitativa realizada junto con los ensayos aleatorizados.
3. Ampliar las revisiones al realizar búsquedas y síntesis para investigar específicamente la evidencia de los estudios cualitativos que aborden preguntas directamente relacionadas con la revisión de efectividad.
4. Complementar las revisiones al sintetizar la evidencia cualitativa que aborde preguntas sobre aspectos que no sean la efectividad.

La síntesis cualitativa para ampliar y complementar las revisiones adoptan un enfoque de síntesis multinivel o paralelo, como se discute en la Sección 20.3.2.5. Actualmente no existen plantillas que permitan realizar una revisión Cochrane exclusivamente de evidencia cualitativa.

El área Cochrane de Salud Pública y Promoción de Salud ha producido guías adicionales sobre los tipos de revisiones y preguntas en las que la investigación cualitativa puede ser de valor (ver Capítulo 21). Dichas revisiones están diseñadas para responder las siguientes preguntas: 1) ¿la intervención funciona? (efectividad); 2) ¿por qué funciona o no funciona (incluido cómo funciona)? (factibilidad, lo apropiado de la intervención y su importancia); y 3) ¿cómo perciben la intervención los participantes?

Cuando la investigación cualitativa se utilice para mejorar o ampliar una revisión Cochrane de intervenciones, los métodos para la especificación, la identificación, la evaluación crítica y la síntesis de la investigación cualitativa se deberían describir en un encabezado separado en “Obtención y análisis de los datos” en la sección “Métodos” de la revisión.

20.2.3 Consideración de los estudios cualitativos que se identifican en los ensayos controlados aleatorizados o junto a ellos

Como los “métodos mixtos” evolucionan para evaluar los efectos de las intervenciones complejas como las estrategias de administración de los servicios de salud, cada vez es más probable que los estudios incluidos en las revisiones Cochrane de intervenciones incluyan o se asocien con investigaciones cualitativas, aunque la evidencia proveniente de los estudios cualitativos puede que no se describa en la misma publicación que los ensayos. Por ejemplo, en un ejemplo de revisión que se resume en el Recuadro 20.3.a, cinco de seis ensayos incluidos en la revisión Cochrane de intervenciones tuvieron un componente o un estudio cualitativo incorporado, aunque no todos los datos cualitativos se habían analizado o

publicado. Es importante señalar que este componente cualitativo no siempre se citó en la publicación del ensayo. De hecho, algunos estudios sólo aparecieron después de establecer contacto con el investigador principal del ensayo.

Cuando se considera la investigación cualitativa identificada dentro de los ensayos aleatorizados o junto a ellos, es necesario valorar los siguientes aspectos:

1. Identificación de la evidencia cualitativa: la evidencia cualitativa recuperada mediante una estrategia de búsqueda de un tema diseñada para identificar ensayos no se puede ver como exhaustiva ni representativa. Dicha estrategia de búsqueda no está diseñada con el objetivo de identificar estudios cualitativos y de hecho logra una medida de especificidad al excluir de forma intencionada muchos tipos de investigación cualitativa.

2. Síntesis de la evidencia cualitativa para explorar la experiencia de presentar la enfermedad: si la experiencia de la enfermedad es el centro de interés, entonces las fuentes cualitativas identificadas en la estrategia de búsqueda del ensayo no proporcionarán necesariamente un criterio holístico o exhaustivo. En estos casos se debería considerar o facilitar la síntesis multinivel o paralela (ver Sección 20.3.2.5). De manera ideal un autor trabajaría con un investigador calificado y un documentalista para desarrollar una estrategia de búsqueda cualitativa para identificar otros estudios relevantes.

3. Síntesis cualitativa para explorar aspectos de la implementación de la intervención: si los aspectos con relación a la implementación son el centro del interés entonces la evidencia cualitativa incorporada o asociada con los ensayos sería la más relevante. Es más probable que esta evidencia se genere mediante métodos mixtos de investigación e incluya evidencia cuantitativa y cualitativa. Es necesario dar pasos para identificar todas las fuentes cualitativas asociadas con los ensayos, como la realización de búsquedas adicionales dirigidas y el contacto con el investigador principal del ensayo.

4. Consideración de la evidencia cualitativa dentro de los estudios excluidos de las revisiones Cochrane de intervenciones: en ocasiones un ensayo no cumple los criterios de elegibilidad de una revisión Cochrane de intervenciones (por ejemplo, debido a un riesgo inaceptable de riesgo), pero la investigación cualitativa incorporada o asociada con el ensayo se considera de alta calidad. El principio guía es que si las pruebas cualitativas parecen ser sólidas, las pruebas cualitativas se pueden incorporar a la revisión.

20.2.4 Consideraciones relacionadas con los recursos

El propósito de incorporar evidencia de la investigación cualitativa en una revisión Cochrane tiene inevitablemente muchas consecuencias para los revisores y los Grupos Cochrane de Revisión (GCR). Las limitaciones de recursos pueden determinar el grado en el cual es posible realizar la síntesis cualitativa complementaria que acompaña a las revisiones. Los revisores deberán considerar los siguientes aspectos cuando pretenden incorporar pruebas de la investigación cualitativa a una revisión Cochrane.

- ¿El equipo tiene la pericia apropiada o el acceso a asesoramiento de investigadores de síntesis cualitativa experimentados?
- ¿Será necesario una formación adicional?
- ¿El presupuesto cubre el tiempo y los recursos adicionales necesarios?
- ¿El equipo tiene acceso a bases de datos y revistas apropiadas?
- ¿El equipo tiene acceso a un documentalista familiarizado con los retos específicos de la recuperación de investigaciones cualitativas?

- ¿El GCR responsable de la revisión apoya la incorporación de evidencia cualitativa y tiene los recursos para apoyar la revisión durante el proceso editorial?

20.3 Síntesis de la evidencia cualitativa

20.3.1 Ejemplo de síntesis de evidencia cualitativa para complementar una revisión Cochrane de intervenciones: tratamiento observado directamente y tuberculosis (TB)

Antes de considerar la metodología para la síntesis de la evidencia cualitativa se proporciona un ejemplo resumido en el Recuadro 20.3.a. La revisión completa está publicada en el *Journal of Advanced Nursing* (Noyes 2007). Esta síntesis de la evidencia cualitativa paralela amplía y complementa una revisión Cochrane de intervenciones de tratamiento observado directamente (ingestión supervisada de la medicación) como una intervención para mejorar el cumplimiento de los pacientes de los regímenes para la tuberculosis (TB) (Volmink 2007), que incluyó seis ensayos aleatorizados pero no encontró un efecto estadísticamente significativo del tratamiento observado directamente (TOD) cuando se comparó con las personas que se autoadministraron el tratamiento en su domicilio. La síntesis acompañante de la evidencia cualitativa se centra en experiencias no profesionales y percepciones del tratamiento para la TB con el fin de considerar si las pruebas de estos estudios pudieran ayudar a explicar los resultados de los ensayos controlados aleatorizados y contribuir al desarrollo de políticas para el tratamiento de la TB. Al hacerlo, la síntesis la evidencia cualitativa abordó interrogantes más allá de la revisión Cochrane de intervenciones como lo apropiado del TOD y la forma en la que el mismo se facilitó en la práctica.

Recuadro 20.3.a: Tratamiento directamente observado de la tuberculosis: una síntesis del resumen de la evidencia cualitativa

Antecedentes: el TOD es parte de un paquete de intervenciones reconocido por la Organización Mundial de la Salud (OMS) para mejorar el tratamiento de la TB y el cumplimiento con el tratamiento (Maher 1999). El TOD incluye pedirles a los pacientes con TB que visiten a un trabajador sanitario, o a otra persona encargada, para que reciban y sean observados mientras toman una dosis de la medicación. Una revisión Cochrane de ensayos de intervenciones de TOD mostró evidencias contradictorias sobre los efectos del TOD cuando se comparó con la autoadministración del tratamiento. Para complementar esta revisión se realizó una síntesis de la evidencia cualitativa con respecto a pacientes con, o en riesgo de tener, TB, proveedores de servicios y personas que toman decisiones, para explorar sus experiencias y percepciones de la TB y el tratamiento. Los hallazgos se utilizaron para ayudar a explicar e interpretar la revisión Cochrane de intervenciones y considerar las implicaciones para la investigación, la legislación y la práctica.

Preguntas de revisión: Se abordaron dos preguntas de revisión amplias:

1. ¿Cuáles son los facilitadores y las barreras para acceder y cumplir con el tratamiento de la tuberculosis?
2. ¿Es posible explicar la heterogeneidad de los hallazgos a través de la exploración de los estudios cualitativos o los componentes cualitativos de los estudios incluidos en la revisión de intervenciones?

Método:

Métodos de búsqueda: se realizó una búsqueda sistemática de la bibliografía más amplia en idioma inglés. Se utilizaron los siguientes términos: DOT; DOTS; *Directly observed therapy; Directly observed treatment; supervised swallowing; self-supervis**;

en combinación con TB y tuberculosis. Se experimentó con el uso de filtros metodológicos al incluir términos como "qualitative", pero se encontró que este enfoque no fue útil porque el encabezado de términos MeSH de Medline "Qualitative Research" no se introdujo hasta 2003, e incluso después de 2003 muchos artículos no se identificaban de forma apropiada como cualitativos. Se buscó en MEDLINE, CINAHL, HMIC, Embase, British Nursing Index, International Bibliography of the Social Sciences, Sociological Abstracts, SIGLE, ASSIA, Psych Info, Econ lit, Ovid, Pubmed, la base de datos de estudios de TB de la London School of Hygiene and Tropical Medicine (cortesía del Dr. Simon Lewin), y en Google Scholar. También se examinaron las listas de referencias de los artículos publicados y una red de contactos personales para identificar artículos. Se estableció contacto con los investigadores principales de los seis ensayos aleatorizados incluidos en la revisión Cochrane de intervenciones y se obtuvieron estudios cualitativos relevantes.

Selección y evaluación de los estudios: se utilizó la siguiente definición para seleccionar los estudios: "artículos cuyo foco primario fueron las experiencias o las percepciones de la TB y su tratamiento entre pacientes que presentaban, o estaban en riesgo de presentar, TB y los proveedores de servicios". El estudio tenía que utilizar métodos cualitativos de obtención y análisis de los datos, como estudios únicos o como parte de un estudio más grande de métodos mixtos. Para evaluar las dimensiones metodológicas y teóricas de la calidad del estudio, JN y JP utilizaron de forma independiente dos marcos de trabajo opuestos (Popay 1998a, Critical Appraisal Skills Programme 2006). Los estudios no se excluyeron debido a su calidad, pero los estudios de menor calidad se examinaron para determinar si alteraban el resultado de la síntesis, lo que no ocurrió.

Análisis: se utilizaron técnicas de análisis temático para resumir los datos de 1990 a 2002, y una actualización de la bibliografía hasta diciembre de 2005. Los temas se identificaron al reunir componentes de ideas, experiencias y criterios incorporados en los datos (los temas se construyeron para formar un cuadro exhaustivo de las experiencias colectivas de los participantes). Se utilizó una técnica de resumen narrativo para ayudar a interpretar los resultados de los ensayos.

Hallazgos: se incluyeron 58 artículos derivados de 53 estudios. De la síntesis de 1990 a 2002 surgieron cinco temas, incluidos: circunstancias socioeconómicas; recursos materiales y agencias individuales; modelos explicativos y sistemas de conocimientos en relación a la tuberculosis y su tratamiento; experiencia del estigma y discursos públicos alrededor de la tuberculosis; sanciones, incentivos y apoyo; y la organización social y las relaciones sociales de la atención. De la actualización de 2005 surgieron dos temas adicionales: las barreras creadas por la implementación del programa y el reto para el modelo de que los factores culturalmente determinados son la causa fundamental del fracaso del tratamiento.

Conclusiones: la revisión de intervenciones Cochrane no mostró diferencias significativas entre el TOD y la autosupervisión, lo que indica que el TOD por sí mismo no dio lugar a una mejoría en los resultados de tratamiento. Los seis ensayos aleatorizados probaron ocho variaciones del TOD comparado con la autosupervisión y variaron mucho en cuanto al grado en el cual se ajustaron a las necesidades de los pacientes con TB. Las variantes de TOD difirieron de forma importante en cuanto a quién fue observado, dónde ocurrió la observación y cuán a menudo se realizó la observación. La síntesis de la investigación cualitativa indica que estos elementos del TOD serán fundamentales para determinar cuán efectivo será un tipo determinado de TOD con respecto al aumento de las tasas de curación. La revisión cualitativa también destacó la función clave de factores sociales y económicos y los efectos secundarios físicos de la medicación para influir en la conducta relacionada con la búsqueda de diagnóstico y el cumplimiento del tratamiento. Más específicamente, no es probable que un enfoque de la observación en el que predomine más la supervisión aumente la captación del servicio ni el cumplimiento con la medicación. Los elementos de supervisión pueden ser necesarios en los paquetes de tratamiento, pero cuando el foco primario de la observación directa tuvo una naturaleza de supervisión en lugar de ser de apoyo, la observación fue menos efectiva. La

observación directa de naturaleza de supervisión tiene la repercusión más negativa sobre quienes tuvieron más temor a la revelación, como las mujeres en desventaja que sufrían discriminación de género. Por el contrario, es más probable que los paquetes de tratamiento en los cuales se hizo énfasis en el apoyo centrado en la persona aumenten la captación y el cumplimiento. Las pruebas cualitativas también proporcionaron cierta introspección sobre el tipo de apoyo que los pacientes con TB encontraron más útil. Principalmente, la habilidad del observador para ser un valor añadido depende del observador y del servicio que es capaz de adaptar a las circunstancias individuales ampliamente variables de la persona que es observada (edad, sexo, agencia, ubicación, ingresos, etc.). Debido a la heterogeneidad entre los pacientes con TB, los hallazgos apoyan la necesidad de programas adaptados y centrados en el paciente en lugar de una intervención global única.

20.3.2 Aspectos metodológicos

Introducción

Los principales retos metodológicos de la síntesis de evidencia cualitativa se relacionan con el diseño y la realización de las estrategias de búsqueda, la evaluación de la calidad del estudio y los métodos apropiados para la síntesis.

20.3.2.1 Estrategias de búsqueda

Se han hecho progresos significativos en el análisis de los sistemas de bases de datos indexadas en busca de estudios cualitativos. El Hedges Project de la Universidad de McMaster ha ampliado su cobertura de filtros metodológicos probados empíricamente para incluir filtros de investigación cualitativa para MEDLINE (Wong 2004), CINAHL (Wilczynski 2007), PsycINFO (McKibbon 2006) y EMBASE (Walters 2006). No obstante, es difícil recuperar pruebas de los estudios cualitativos obtenidos y descritos dentro de los ensayos aleatorizados o como parte de estudios vinculados (Evans 2002). MEDLINE apenas introdujo el término MeSH “qualitative research’ en 2003. CINAHL introdujo “Qualitative Studies” en 1988, lo que refleja el interés específico en los estudios cualitativos para los investigadores de enfermería, con un foco correspondiente sobre aspectos de la “calidad de vida” (ver Capítulo 17, Sección 17.3). Sin embargo, la localización de estudios cualitativos aún es problemática debido al uso variable del término “cualitativa” (Grant 2004).

Además, las estrategias actuales para indexar términos relacionados con los diseños de estudios cualitativos y las estrategias de búsquedas guiadas por protocolos son de valor limitado (Evans 2002, Barroso 2003, Greenhalgh 2005). Los revisores deberían saber que limitar una búsqueda a bases de datos bien conocidas puede dar lugar a la pérdida de mucha información útil. Una auditoría de las fuentes para una revisión de intervenciones complejas (incluidas las pruebas cualitativas) encontró que sólo el 30% se identificó a partir de bases de datos y búsquedas manuales. Cerca de la mitad de los estudios se identificó mediante la “bola de nieve” y otro 24% mediante el conocimiento o el contacto personales (Greenhalgh 2005). Las estrategias de búsqueda para identificar estudios cualitativos que utilicen varios métodos cualitativos se debería desarrollar aún más.

Aunque existe acuerdo general sobre la necesidad de que las estrategias de búsqueda dirigidas a identificar investigaciones cualitativas sean sistemáticas y explícitas, ha habido un debate reciente acerca de si la síntesis de la evidencia cualitativa comparte la necesidad de ser exhaustiva. Se ha argumentado que puede ser más apropiado un enfoque de muestra más decidido, dirigido a proporcionar una interpretación holística de un fenómeno, en el que el grado de búsqueda esté guiado por la necesidad de alcanzar la saturación

teórica y la identificación del "caso no confirmatorio" (Dixon-Woods 2006). No obstante, esto hace aún más imperativo mejorar la calidad de los estándares de publicación de los métodos de búsqueda (Booth 2006).

20.3.2.2 Evaluación crítica

La evaluación de la calidad de un estudio (evaluación crítica) es un aspecto particularmente debatido con respecto a la síntesis de las pruebas cualitativas. Hasta el momento la opinión sobre el valor de la evaluación formal de la calidad está dividida y no hay pruebas suficientes para informar una valoración sobre el rigor o el valor agregado de los diferentes enfoques.

Éste es un campo en evolución y los miembros del Grupo Cochrane de Métodos de Investigaciones Cualitativas está activamente involucrado en contribuir al conocimiento y la práctica en esta área. Sin embargo, se considera que es importante analizar y debatir los argumentos a favor y en contra de la evaluación crítica en la síntesis de la evidencia cualitativa.

Cerca de cien herramientas y enfoques están disponibles para ayudar a la evaluación de la investigación cualitativa, los cuales reflejan los que están disponibles para la evaluación de la calidad metodológica de los ensayos aleatorizados y otras formas de investigación cuantitativa (Vermeire 2002, Cote 2005). Sin embargo, es importante reconocer que las interrogantes acerca de la "calidad" son muy diferentes en el contexto de la investigación cualitativa. Se puede argumentar que los procesos de evaluación formal y los estándares de la evidencia presentada como listas de verificación rígidas que describen una decisión "dentro o fuera" no son apropiados para el contexto de la investigación cualitativa (Popay 1998a, Barbour 2001, Spencer 2003). Por el contrario, dichas herramientas quizás se utilicen mejor como parte de un proceso de exploración e interpretación. Los estudios calificados de calidad deficiente según un método tipo formulario rígido pueden generar nuevas introspecciones basadas en los datos, mientras que los estudios metodológicamente sólidos pueden tener una interpretación deficiente y dar lugar a una introspección insuficiente del fenómeno en estudio. Dixon-Woods y cols. compararon tres enfoques de evaluación estructurados y concluyeron que los enfoques estructurados pueden no producir valoraciones más consistentes acerca de si incluir los artículos cualitativos en una revisión sistemática (Dixon-Woods 2007).

Un aspecto adicional se relaciona con el momento de la evaluación de la calidad y con cuándo se deberían tener en cuenta los resultados del proceso: ¿la evaluación crítica se debería considerar una barrera para establecer un umbral de calidad o como un filtro para mediar las diferentes fortalezas de los mensajes resultantes de la investigación incluida?

Si los autores deciden incorporar la evaluación de la calidad como parte del proceso de revisión sistemática entonces pueden utilizar el enfoque integral para el método específico (como el enfoque de Evidencia para la Política y la Información de la Práctica [EPPI por sus siglas en inglés] o el enfoque del Joanna Briggs Institute [JBI]), o seleccionar cualquier herramienta, marco de trabajo o lista de verificación de evaluación cualitativa publicadas. Spencer y cols. realizaron una revisión de muchos de los marcos de trabajo y listas de verificación de evaluación actuales, la cual los revisores pueden encontrar útiles para decidir qué enfoque aplicar (Spencer 2003). La valoración de los expertos también es un factor importante cuando se evalúa la calidad de los estudios.

Las referencias clave que reflejan este debate se incluyen en la Sección 20.6.6: Lecturas adicionales.

20.3.2.3 Síntesis de la evidencia de la investigación cualitativa

La síntesis de las pruebas cualitativas es un proceso en el que se combinan las pruebas de estudios cualitativos individuales para crear una nueva comprensión al comparar y analizar conceptos y hallazgos de diferentes fuentes de evidencia enfocadas en el mismo tema de interés. Por lo tanto, la síntesis de la evidencia cualitativa se puede considerar como un estudio completo en sí mismo, comparable a cualquier metanálisis dentro de una revisión sistemática sobre efectos de intervenciones o pruebas diagnósticas. La misma puede ser un proceso de agregación o interpretativo pero requiere transparencia del proceso y que los revisores identifiquen y extraigan las pruebas de los estudios incluidos en la revisión, categoricen las pruebas y combinen estas categorías para desarrollar hallazgos resumidos. Sin embargo, si se realiza este trabajo metodológico es importante reconocer que el valor real de la síntesis de las pruebas cualitativas no es sólo una descripción de cómo se sienten las personas acerca de un problema o un tratamiento, sino la comprensión de "por qué" se sienten y actúan como lo hacen (Popay 2005).

Por ejemplo, la investigación cualitativa primaria sobre la experiencia de la enfermedad crónica presenta la estimación de las personas sobre el inicio de la enfermedad. No obstante, este trabajo también va más allá de la descripción y busca explicar el propósito social de estas estimaciones (muestra cómo a través de estas narraciones las personas "reconstruyen" un sentido de valor en un contexto social en el cual todas las enfermedades tienen trasfondos morales) (Williams 1984). De manera similar, una revisión sistemática reciente de investigación cualitativa sobre la ingestión de fármacos (Campbell 2003, Pound 2005) que utilizó la metaetnografía como un método para la síntesis, fue más allá y brindó un resumen de "temas" recurrentes entre los estudios para construir una explicación de por qué las personas utilizan las medicaciones (o no) de la forma que lo hacen.

20.3.2.4 Selección del método apropiado

La selección del método apropiado para la inclusión de evidencia cualitativa en una síntesis de evidencia cualitativa dependerá de varios factores, entre ellos:

- tipo y alcance de la revisión y la/s pregunta/s de revisión;
- grupo de evidencia disponible;
- pericia del equipo; y
- recursos disponibles.

Hay varios métodos en evolución para la síntesis de evidencia cualitativa y evidencia de métodos mixtos. Junto con otros individuos y organizaciones de revisiones sistemáticas, el grupo Cochrane de Métodos de Investigación Cualitativa está activamente involucrado en desarrollar y más recientemente comenzar a evaluar el rango de métodos disponibles. Los miembros han contribuido a dos textos fundamentales sobre la síntesis de la evidencia cualitativa y cuantitativas en salud, lo que brinda información más detallada y guías sobre los métodos y procesos (Petticrew 2006, Pope 2007).

Se recomienda que se utilice cualquier método de alta calidad para la síntesis de las pruebas cualitativas que se adapte mejor al tipo de revisión de intervenciones.

Está más allá del alcance de este capítulo incluir información detallada del rango de métodos disponibles para la síntesis de pruebas cualitativas y de métodos mixtos. Se han utilizado varios métodos en las revisiones publicadas. Los ejemplos incluyen: metanálisis bayesiano, síntesis crítica interpretativa, enfoque del Centro de Evidencia para la Política y la Información de la Práctica (EPPI), el enfoque del *Joanna Briggs Institute* (JBI), la

metaetnografía, la metasíntesis, el metaestudio, el metarresumen, la síntesis narrativa, la síntesis de las pruebas cualitativas basada en la teoría de base (*grounded theory*), la síntesis realista y el análisis temático secundario.

La mayoría de los métodos tienen guías detalladas asociadas (ver por ejemplo Noblit y Hare para la metaetnografía y Popay y cols. para la síntesis narrativa (Noblit 1988, Popay 2006b)), las cuales se deberían citar. Dixon-Woods y cols. brindan un resumen detallado del potencial de varios métodos y los retos asociados (Dixon-Woods 2005, Dixon-Woods 2006). Hasta el momento se han realizado pocas investigaciones para determinar la solidez de los diferentes métodos. Se pueden encontrar lecturas adicionales en la Sección 20.6.

20.3.2.5 Enfoques para integrar la síntesis de la evidencia cualitativa y cuantitativa

Existen dos enfoques amplios que se pueden utilizar para integrar los hallazgos cualitativos y cuantitativos:

1. Síntesis multinivel: las síntesis de la evidencia cualitativa (síntesis 1) y la evidencia cuantitativa (síntesis 2) se pueden realizar como grupos separados o revisiones separadas pero vinculadas y el producto de cada síntesis se combina luego (síntesis 3) (ver, por ejemplo, Thomas y cols. (Thomas 2004)).
2. Síntesis paralela: las síntesis de la evidencia cualitativa (síntesis 1) y la evidencia cuantitativa (síntesis 2) se pueden realizar como grupos separados o revisiones separadas pero vinculadas. La síntesis cualitativa (síntesis 1) se puede utilizar luego en paralelo o yuxtapuesta para ayudar a la interpretación de los ensayos resumidos (síntesis 2) (ver, por ejemplo, Noyes y Popay (Noyes 2007)).

La síntesis multinivel y la paralela requieren una revisión sistemática separada de la evidencia, que en un estadio posterior se resume o se yuxtapone a la síntesis de los ensayos. Las pautas sobre la realización de síntesis narrativas (Popay 2006b) contienen una herramienta para reunir los hallazgos de los diferentes diseños de estudios dentro de los diferentes métodos y enfoques. Se requiere un trabajo metodológico adicional sobre los procesos por los cuales la evidencia de los estudios que utilizan diferentes métodos cualitativos y generan un rango de tipos de evidencia se pueden resumir y combinar con los hallazgos cuantitativos sobre el efecto sin comprometer la necesidad de minimizar el sesgo (Lucas 2007).

20.3.2.6 Conclusión

El interés en revisar sistemáticamente formas más amplias de evidencia y específicamente, evidencia proveniente de investigación cualitativa, se basa en el creciente reconocimiento de que la investigación cualitativa puede mejorar la relevancia y utilidad de una revisión. Sin embargo, la investigación que se genera de forma rigurosa, independientemente del diseño, exige una evaluación cuidadosa de su calidad antes de que se pueda utilizar en el medio clínico. Para que se tenga en cuenta para una revisión Cochrane de intervenciones, la evidencia de la investigación cualitativa debería estar sujeta a métodos de revisión igualmente rigurosos. Actualmente surgen métodos de evaluación y análisis de la evidencia de la investigación cualitativa que evoluciona con el tiempo. Se necesita evidencia adicional para establecer el rigor y el valor agregado de varios enfoques para la evaluación de la calidad en el proceso de revisión sistemática.

20.4 Información del capítulo

Autores: Jane Noyes, Jennie Popay, Alan Pearson, Karin Hannes y Andrew Booth del Grupo Cochrane de Métodos de Investigación Cualitativa.

La versión en inglés de este capítulo se debería citar como: Noyes J, Popay J, Pearson A, Hannes K, Booth A. Chapter 20: Qualitative research and Cochrane reviews. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Recuadro 20.4.a: El Grupo Cochrane de Métodos de Investigación Cualitativa

El Grupo Cochrane de Métodos de Investigación Cualitativa (GCMIC) desarrolla y apoya el trabajo metodológico sobre la inclusión en las revisiones sistemáticas de la evidencia de las investigaciones que utilizan métodos cualitativos y difunde este trabajo dentro y más allá de los GCR de la Colaboración.

El GCMIC intenta cumplir esta función al:

- Identificar funciones apropiadas para la evidencia proveniente de las investigación cualitativa dentro del contexto de las revisiones sistemáticas Cochrane.
- Cotejar, desarrollar y difundir estándares metodológicos apropiados para la:
    a. búsqueda de investigaciones cualitativas relevantes para las revisiones Cochrane;
    b. evaluación crítica de los estudios cualitativos;
    c. combinación de la evidencia de la investigación cualitativa con otros datos dentro del contexto de una revisión sistemática; y
    d. difusión de estos estándares metodológicos de diferentes formas incluida la contribución a las guías para los revisores en el *Manual*.
- Proporcionar un foro para la discusión y el debate acerca de la función de la evidencia cualitativa dentro del proceso de revisión sistemática y el desarrollo de métodos rigurosos y sistemáticos para promover dicha función para:
    e. estimular la transparencia y el aprendizaje del desarrollo de métodos; y
    f. estimular y facilitar el vínculo y la participación conjunta con otros grupos de métodos.
- Proporcionar vínculos a los Grupos Cochrane de Revisión a personas con pericia y experiencia en investigación cualitativa para:
    g. proporcionar asesoramiento y apoyo a las personas que pretendan incorporar la investigación cualitativa a una revisión; y
    h. proporcionar un mecanismo para evaluar y desarrollar protocolos de revisión.
- Proporcionar formación a miembros de los Grupos Cochrane y Campbell de Revisión.
- Mantener un registro/base de datos de artículos metodológicamente relevantes.
- Mantener un registro/base de datos de protocolos de revisiones sistemáticas que incluyen síntesis de pruebas cualitativas o que se centran únicamente en la revisión sistemática de evidencia cualitativa.
- Mantener un registro/base de datos de revisiones sistemáticas completadas que incluyen síntesis de evidencia cualitativa, y de revisiones que se centran únicamente en la revisión sistemática de evidencia cualitativa.
- Encuestar anualmente a los miembros para identificar los intereses de desarrollo y las contribuciones en marcha.

Los miembros del Grupo han contribuido a guiar la realización de revisiones sistemáticas producidas por el Centre for Reviews and Dissemination en la Universidad de Cork y han apoyado el desarrollo de guías producidas por Área Cochrane de Promoción de Salud y Salud Pública.

Web site: www.joannabriggs.edu.au/cqrmg

20.5 Referencias

Arai 2005
Arai L, Roen K, Roberts H, Popay J. It might work in Oklahoma but will it work in Oakhampton? Context and implementation in the effectiveness literature on domestic smoke detectors. *Injury Prevention* 2005; 11: 148-151.

Barbour 2001
Barbour RS. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *BMJ* 2001; 322: 1115-1117.

Barroso 2003
Barroso J, Gollop CJ, Sandelowski M, Meynell J, Pearce PF, Collins LJ. The challenges of searching for and retrieving qualitative studies. *Western Journal of Nursing Research* 2003; 25: 153-178.

Booth 2006
Booth A. "Brimful of STARLITE": toward standards for reporting literature searches. *Journal of the Medical Library Association* 2006; 94: 421-429.

Campbell 2003
Campbell R, Pound P, Pope C, Britten N, Pill R, Morgan M, Donovan J. Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. *Social Science and Medicine* 2003; 56: 671-684.

Cote 2005
Cote L, Turgeon J. Appraising qualitative research articles in medicine and medical education. *Medical Teacher* 2005; 27: 71-75.

Critical Appraisal Skills Programme 2006
Critical Appraisal Skills Programme. 10 questions to help you make sense of qualitative research [2006]. Available from: http://www.phru.nhs.uk/Pages/PHD/resources.htm (accessed 1 January 2008).

Denzin 1994
Denzin NK, Lincoln YS. Introduction. Entering the field of qualitative research. In: Denzin NK, Lincoln YS (editors). *Handbook of Qualitative Research*. Thousand Oaks (CA): Sage Publications, 1994.

Dixon-Woods 2005
Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. *Journal of Health Services Research and Policy* 2005; 10: 45-53.

Dixon-Woods 2006
Dixon-Woods M, Bonas S, Booth A, Jones DR, Miller T, Sutton AJ, Shaw RL, Smith JA, Young B. How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research* 2006; 6: 27-44.

Dixon-Woods 2007
Dixon-Woods M, Sutton A, Shaw R, Miller T, Smith J, Young B, Bonas S, Booth A, Jones D. Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. *Journal of Health Services Research and Policy* 2007; 12: 42-47.

Evans 2002
Evans D. Database searches for qualitative research. *Journal of the Medical Library Association* 2002; 90: 290-293.

Grant 2004
Grant MJ. How does your searching grow? A survey of search preferences and the use of optimal search strategies in the identification of qualitative research. *Health Information and Libraries Journal* 2004; 21: 21-32.

Greenhalgh 2005
Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ* 2005; 331: 1064-1065.

Lucas 2007
Lucas PJ, Baird J, Arai L, Law C, Roberts HM. Worked examples of alternative methods for the synthesis of qualitative and quantitative research in systematic reviews. *BMC Medical Research Methodology* 2007; 7: 4.

Maher 1999
Maher D, Mikulencak M. *What is DOTS? A Guide to Understanding the WHO-recommended TB Control Strategy Known as DOTS*. Geneva (Switzerland): World Health Organization, 1999.

Mays 2005
Mays N, Pope C, Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *Journal of Health Services Research and Policy* 2005; 10 (Suppl 1): 6-20.

McKibbon 2006
McKibbon KA, Wilczynski NL, Haynes RB. Developing optimal search strategies for retrieving qualitative studies in PsycINFO. *Evaluation and the Health Professions* 2006; 29: 440-454.

Mills 2005a
Mills E, Jadad AR, Ross C, Wilson K. Systematic review of qualitative studies exploring parental beliefs and attitudes toward childhood vaccination identifies common barriers to vaccination. *Journal of Clinical Epidemiology* 2005; 58: 1081-1088.

Mills 2005b
Mills EJ, Montori VM, Ross CP, Shea B, Wilson K, Guyatt GH. Systematically reviewing qualitative studies complements survey design: an exploratory study of barriers to paediatric immunisations. *Journal of Clinical Epidemiology* 2005; 58: 1101-1108.

Noblit 1988
Noblit GW, Hare RD. *Meta-ethnography: Synthesising Qualitative Studies* (Qualitative Research Methods). London: Sage Publications, 1988.

Noyes 2007
Noyes J, Popay J. Directly observed therapy and tuberculosis: how can a systematic review of qualitative research contribute to improving services? A qualitative meta-synthesis. *Journal of Advanced Nursing* 2007; 57: 227-243.

Pearson 2005
Pearson A, Wiechula R, Court A, Lockwood C. The JBI model of evidence-based healthcare. *JBI Reports* 2005; 3: 207-216.

Petticrew 2006
Petticrew M, Roberts H. *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford (UK): Blackwell, 2006.

Popay 1998a
Popay J, Rogers A, Williams G. Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research* 1009; 8: 341-351.

Popay 1998b
Popay J, Williams G. Qualitative research and evidence-based healthcare. *Journal of the Royal Society of Medicine* 1998; 91 (Suppl 35): 32-37.

Popay 2005
Popay J. Moving beyond floccinaucinihilipilification: enhancing the utility of systematic reviews. *Journal of Clinical Epidemiology* 2005; 58: 1079-1080.

Popay 2006a
Popay J. Incorporating qualitative information in systematic reviews. *14th Cochrane Colloquium*, Dublin (Ireland), 2006.

Popay 2006b
Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, Britten N, Roen K, Duffy S. Guidance on the conduct of narrative synthesis in systematic reviews. Results of an ESRC funded research project. (Unpublished report, 2006, University of Lancaster, UK).

Pope 2006
Pope C, Mays N. Qualiative methods in health research. In: Pope C, Mays N (editors). *Qualitative Research in Health Care* (3rd edition). Malden (MA): Blackwell Publications/BMJ Books, 2006.

Pope 2007
Pope C, Mays N, Popay J. *Synthesising Qualitative and Quantitative Health Research: A Guide to Methods*. Maidenhead (UK): Open University Press., 2007.

Pound 2005
Pound P, Britten N, Morgan M, Yardley L, Pope C, Daker-White G, Campbell R. Resisting medicines: a synthesis of qualitative studies of medicine taking. *Social Science and Medicine* 2005; 61: 133-155.

Roen 2006
Roen K, Arai L, Roberts H, Popay J. Extending systematic reviews to include evidence on implementation: methodological work on a review of community-based initiatives to prevent injuries. *Social Science and Medicine* 2006; 63: 1060-1071.

Spencer 2003
Spencer L. *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence*. London (UK): Government Chief Social Researcher's Office, Cabinet Office, 2003. Available from www.gsr.gov.uk/downloads/evaluating_policy/a_quality_framework.pdf.

Thomas 2004
Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R, Brunton G, Kavanagh J. Integrating qualitative research with trials in systematic reviews. *BMJ* 2004; 328: 1010-1012.

Vermeire 2002
Vermeire E, Van Royen P, Griffiths F, Coenen S, Peremans L, Hendrickx K. The critical appraisal of focus group research articles. *European Journal of General Practice* 2002; 8: 104-108.

Volmink 2007
Volmink J, Garner P. Directly observed therapy for treating tuberculosis. *Cochrane Database of Systematic Reviews* 2006, Issue 4. Art No: CD003343.

Walters 2006
Walters LA, Wilczynski NL, Haynes RB. Developing optimal search strategies for retrieving clinically relevant qualitative studies in EMBASE. *Qualitative Health Research* 2006; 16: 162-168.

Wilczynski 2007
Wilczynski NL, Marks S, Haynes RB. Search strategies for identifying qualitative studies in CINAHL. *Qualitative Health Research* 2007; 17: 705-710.

Williams 1984
Williams G. The genesis of chronic illness: narrative re-construction. *Sociology of Health and Illness* 1984; 6: 175-200.

Wong 2004
Wong SS, Wilczynski NL, Haynes RB, Hedges Team. Developing optimal search strategies for detecting clinically relevant qualitative studies in MEDLINE. *Medinfo* 2004; 11: 311-316.

20.6 Further selected reading

20.6.1 Qualitative research, general

Boulton M, Fitzpatrick R. Qualitative methods for assessing health care. *Quality in Health Care* 1994; 3: 107-113.
Britten N, Jones R, Murphy E, Stacey R. Qualitative research methods in general practice and primary care. *Family Practice* 1995; 12:104-114
Esterberg KG. *Qualitative Methods in Social Research*. Boston (US): McGraw-Hill, 2002.
Giacomini MK. The rocky road: qualitative research as evidence. *Evidence-Based Medicine* 2001; 6: 4-5
Grbich C. *Qualitative Research in Health: An Introduction*. London (UK): Sage Publications, 1999.
Green J, Britten N. Qualitative research and evidence-based medicine. *BMJ* 1998; 316:1230-2.
Guba RG, Lincoln YS. Competing paradigms in qualitative research. In: Denzin NK, Lincoln YS (Eds) *Handbook of Qualitative Research*. Thousand Oaks (CA): Sage Publications, 1994.
Miller S, Fredericks M. The nature of “evidence” in qualitative research methods. *International Journal of Qualitative Methods* 2003; 2: Article 4. Retrieved 1 January 2008 from http://www.ualberta.ca/~ijqm.
Murphy E, Dingwall R, Greatbach D, Parker S, Watson P. Qualitative research methods in health technology assessment: a review of the literature. *Health Technology Assessment* 1998; 2: 1–274.
Popay J, Williams G. Qualitative research and evidence based healthcare. *Journal of the Royal Society of Medicine* 1998; 91(Suppl 35):32–37.

Pope C, Mays N. Qualitative research: reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health service research. *BMJ* 1995; 311: 42-45.
Pope C, Van Royen P, Baker R. Qualitative methods in research on healthcare quality. *Quality and Safety in Health Care* 2002; 11:148-152.

20.6.2 Qualitative methods

Fetterman DM. *Ethnography. Step by step*. Newbury Park (CA): Sage Publications, 1989.
Glaser BG, Strauss AL. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago (IL): Aldine, 1967.
Hammersley M. *Reading Ethnographic Research*. New York (NY): Langman, 1990.
Hammersley M, Atkinson P. *Ethnography: Principles in Practice*. London (UK): Routledge, 1995 .
Lambert H, McKevitt C. Anthropology in health research: from qualitative methods to multidisciplinarity. *BMJ* 2002; 325: 210-213.
Maggs-Rapport F. Combining methodological approaches in research: ethnography and interpretive phenomenology. *Journal of Advanced Nursing* 2000; 31: 219-225.
Meyer J. Using qualitative methods in health related action research. In: Pope C, Mays N (Eds). *Qualitative Research in Health Care*. London (UK): BMJ Books, 1999.
Savage J. Ethnography and health care. *BMJ* 2000; 321:1400-1402.
Strauss A, Corbin J. *Grounded Theory in Practice*. Thousand Oaks (CA): Sage Publications, 1997.
Strauss A, Corbin J. *Basics of Qualitative Research Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks (CA): Sage Publications, 1998.
Taylor SJ, Bogdan R. *Introduction to Qualitative Research Methods: A Guidebook and Resource*. New York (NY), John Wiley & Sons, 1998.
Yin RK. *Case Study Research: Designs and Methods*. Newbury Park (CA): Sage Publications, 1989.

20.6.3 Qualitative literature searching

Flemming K, Briggs M. Electronic searching to locate qualitative research: evaluation of three strategies. *Journal of Advanced Nursing* 2007; 57: 95-100.
Shaw RL, Booth A, Sutton AJ, Miller T, Smith JA, Young B, Jones DR, Dixon-Woods M. Finding qualitative research: an evaluation of search strategies. *BMC Medical Research Methodology* 2004; 4: 5
InterTASC Information Subgroup, University of York web site:
i. http://www.york.ac.uk/inst/crd/intertasc/

20.6.4 Synthesizing qualitative evidence

Jensen LA, Allen MN. Meta-synthesis of qualitative findings. *Qualitative Health Research* 1996; 6: 553-560.
Noblit GW, Hare RD. *Meta-Ethnography: Synthesising Qualitative Studies*. Newbury Park (CA): Sage Publications, 1988.
Paterson BL, Thorne SE, Canam C, Jillings C. *Meta-Study of Qualitative Health Research. A Practical Guide to Meta-Analysis and Meta-Synthesis*. Thousand Oaks (CA): Sage Publications, 2001.
Pearson A. Balancing the evidence: incorporating the synthesis of qualitative data into systematic reviews. *JBI Reports* 2004; 2 :45-64.
Sandelowski M, Barroso. Creating metasummaries of qualitative findings. *Nursing Research* 2003; 52: 226-33.

Sandelowski M, Barroso J. *Handbook for Synthesising Qualitative Research*. New York (NY): Springer, 2007.
Sandelowski M, Docherty S, Emden C. Focus on qualitative methods. Qualitative meta-synthesis: issues and techniques. *Research in Nursing and Health* 1997; 20: 365-371.
Thorne S, Jensen L, Kearney MH, Noblit G, Sandelowski M. Qualitative metasynthesis: reflections on methodological orientation and ideological agenda. *Qualitative Health Research* 2004; 14: 1342-1365.
Zhao S. Metatheory, metamethod, qualitative meta-analysis: what, why and how? *Sociological Perspectives* 1991; 34: 377-390.

20.6.5 Synthesizing qualitative and quantitative evidence

Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J, Hsu R, Katbamna S, Olsen R, Smith L, Riley R, Sutton AJ. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology* 2006; 6: 35.
Dixon-Woods M, Fitzpatrick R, Roberts K. Including qualitative research in systematic reviews; opportunities and problems. *Journal of Evaluation in Clinical Practice* 2001; 7: 125-133.
Dixon-Woods M, Fitzpatrick R. Qualitative research in systematic reviews. *BMJ* 2001; 323: 765-766
Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, Peacock R. Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. *Social Science and Medicine* 2005; 61: 417-430.
Harden A, Garcia J, Oliver S, Rees R, Shepherd J, Brunton G, Oakley A. Applying systematic review methods to studies of people's views: an example from public health research. *Journal of Epidemiology and Community Health* 2004; 58: 794-800.
Pawson, R. Evidence-based policy: the promise of 'realist synthesis'. *Evaluation* 2002; 8: 340-358.
Pawson R. *Evidence Based Policy: A Realist Perspective*. London (UK): Sage Publications, 2006.
Pearson, A, Field, J, Jordan, Z. *Evidence-based Clinical Practice in Nursing and Healthcare: Assimilating Research, Experience and Expertise*. Oxford (UK): Blackwell, 2007.
Petticrew M, Roberts H. *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford (UK): Blackwell, 2006.
Pope C, Mays N, Popay J. *Synthesising Qualitative and Quantitative Health Research: A Guide to Methods*. Maidenhead (UK): Open University Press, 2007.
Popay J (Ed). *Moving beyond Effectiveness in Evidence Synthesis: Methodological Issues in the Synthesis of Diverse Sources of Evidence*. London (UK): NICE, 2006.
Roberts K, Dixon-Woods M, Fitzpatrick R, Abrams K, Jones D. Factors affecting uptake of childhood immunisation: a Bayesian synthesis of qualitative and quantitative evidence. *The Lancet* 2002; 360: 1596-1599.
Webb C, Roe B (Eds). *Reviewing Research Evidence for Nursing Practice*. Oxford (UK): Blackwell, 2007.

20.6.6 Critical appraisal of qualitative studies

Blaxter M. Criteria for evaluation of qualitative research. *Medical Sociology News* 1996; 22: 68-71.
CASP (Critical Appraisal Skills Programme). 10 Questions to make sense of qualitative research [2006]. Available from: http://www.phru.nhs.uk/pages/phd/resources.htm (accessed 1 January 2008).

Dixon-Woods M, Shaw RL, Agarwal S, Smith JA. The problem of appraising qualitative research. *Quality and Safety in Healthcare* 2004; 13: 223-225.
Elder NC, Miller WL. Reading and evaluation qualitative research studies. *Journal of Family Practice* 1995; 41: 279-285
Forchuk C, Roberts J. How to critique qualitative research articles. *Canadian Journal of Nursing Research* 1993; 25: 47-55.
Horsburgh D. Evaluation of qualitative research. *Journal of Clinical Nursing* 2003; 12: 307-312.
Malterud K Qualitative research: standards, challenges, and guidelines. *The Lancet* 2001; 358: 483-488.
Popay J, Rogers A, Williams G. Rationale and standards for the systematic review of qualitative literature in health service research. *Qualitative Health Research* 1998; 8: 341-351.
Secker J, Wimbush E, Watson J, Milburn K. Qualitative methods in health promotion research: some criteria for quality. *Health Education Journal* 1995; 54: 74-87.
Spencer L, Ritchie J, Lewis J, Dillon L. *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence*. London (UK): Government Chief Social Researcher's Office, 2003.
Vermeire E, Van Royen P, Griffiths F, Coenen S, Peremans L, Hendrickx K. The critical appraisal of focus group research articles. *European Journal of General Practice* 2002; 8: 104-108.

20.6.7 Web sites

(Accessed 1 January 2008)

Campbell Collaboration
A Campbell Review can include evidence from studies of the implementation of an intervention.

j. www.campbellcollaboration.org

Centre for Reviews and Dissemination (CRD), University of York, UK
In addition to a handbook, CRD has an online resource centre.

k. www.york.ac.uk/inst/crd

Evidence for Policy and Practice Information and Coordinating (EPPI) Centre
The EPPI Centre provides links to methods, tools and databases.

l. eppi.ioe.ac.uk/cms

Joanna Briggs Institute (JBI)
JBI offers a variety of evidence-based healthcare resources concerning the synthesis of evidence.

m. www.joannabriggs.edu.au

National Institute for Health and Clinical Excellence (NICE)
NICE has produced guidance on methods for development of NICE public health guidance which incorporate diverse study designs.

n. www.nice.org.uk

Social Care Institute for Excellence (SCIE)
SCIE has produced guidance on the conduct of knowledge reviews which incorporate diverse study designs.

o. www.scie.org.uk

ÍNDICE

CAPÍTULO 21: REVISIONES EN SALUD PÚBLICA Y PROMOCIÓN DE SALUD

PUNTOS CLAVE

21.1 INTRODUCCIÓN

21.2 DISEÑOS DE ESTUDIO A INCLUIR

21.3 BÚSQUEDA

Tabla 21.3.a: Bases de datos relevantes para salud pública y promoción de salud (sitios web enumerados por bases de datos libremente disponibles vía Internet)

21.4 EVALUACIÓN DE LA CALIDAD DEL ESTUDIO Y EL RIESGO DE SESGO

21.5 ÉTICA Y DESIGUALDADES

21.6 CONTEXTO

Figura 21.6.a: Ejemplo de éxito de una intervención según el contexto en el cual se implementó (Frommer 2003)

21.7 SOSTENIBILIDAD

21.8 APLICABILIDAD Y TRANSFERIBILIDAD

21.9 INFORMACIÓN DEL CAPÍTULO

21.10 REFERENCIAS

CAPÍTULO 21: REVISIONES EN SALUD PÚBLICA Y PROMOCIÓN DE SALUD

Editores: Rebecca Armstrong, Elizabeth Waters y Jodie Doyle.

Puntos clave

- Las intervenciones en salud pública y promoción de salud son actividades ampliamente definidas que se evalúan mediante una amplia variedad de enfoques y diseños de estudios que incluyen los ensayos aleatorizados por grupos. En algunos aspectos la mejor evidencia disponible puede provenir de estudios no aleatorizados.
- La búsqueda de bibliografía de salud pública y promoción de salud puede ser una tarea muy compleja y requiere que los revisores utilicen otros métodos además de la búsqueda en las bases de datos para recuperar los estudios.
- Las revisiones sistemáticas de intervenciones en salud pública y promoción de salud tienen la posibilidad de investigar resultados diferenciales para grupos con diferentes niveles de desventajas. Sin embargo, abordar las desigualdades es complicado no sólo por lo limitado de la obtención de información acerca de las diferencias entre los grupos, sino también porque existe una participación limitada de los grupos con desventajas en la investigación.
- Un problema adicional al revisar las intervenciones en salud pública y promoción de salud es cómo separar los efectos de la intervención de la influencia del contexto en el cual se implementó la intervención.
- Se debería solicitar información sobre los factores contextuales y sobre las características de la intervención que puedan explicar el grado en el cual la intervención o los resultados son sustentables.

21.1 Introducción

En 2005 el Área Cochrane de Promoción de Salud y Salud Pública (en inglés *HPPH*) (que ha pasado a ser ahora el Grupo de Revisiones Cochrane de Salud Pública) desarrolló guías específicas para realizar revisiones de intervenciones en salud pública y promoción de salud que se actualizaron en 2007. Este capítulo proporciona una visión general de aspectos específicos para la promoción de salud y la salud pública que no se discuten en otras partes del *Manual*. A la versión completa de las *Guías para Revisiones Sistemáticas de Promoción de Salud y Salud Pública* se puede acceder en el sitio web del Grupo de Revisiones Cochrane de Salud Pública: www.ph.cochrane.org.

21.2 Diseños de estudio a incluir

La salud pública y la promoción de salud con actividades ampliamente definidas que se evalúan mediante una amplia variedad de enfoques y diseños. No se puede utilizar un método único para responder todos los aspectos relevantes relacionados con los problemas e intervenciones en salud pública y promoción de salud. Si la pregunta de revisión se ha especificado claramente entonces se deberían seguir automáticamente los tipos de diseño de estudio necesarios para responderla (Petticrew 2003). Una búsqueda preliminar también ayudaría a identificar los tipos de diseño de estudio que se pudieran utilizar para estudiar la intervención. Los criterios utilizados para seleccionar los estudios deberían reflejar principalmente la pregunta o preguntas que se responderán en la revisión, en lugar de alguna jerarquía predeterminada (Glasziou 2004). Las decisiones acerca de qué tipo/s de diseño/s de estudio incluir influirá en las fases posteriores de la revisión, especialmente en la búsqueda, la evaluación del riesgo de sesgo y el análisis (especialmente para los metanálisis).

Los ensayos aleatorizados proporcionan una fuente útil de evidencia de efectividad, aunque sus resultados pueden tener una generalizabilidad limitada (Black 1996). Para muchas de las intervenciones en promoción de salud y salud pública no hay ensayos aleatorizados disponibles debido a aspectos que incluyen la factibilidad y la ética. El campo de la salud pública ha aumentado la adopción de ensayos aleatorizados por grupos; algunas intervenciones necesitan de su aplicación a nivel de grupo (Donner 2004). Estos ensayos pueden contribuir con evidencia valiosa si se asigna al azar un número suficiente de unidades para asegurar incluso la distribución de factores de confusión potenciales entre los grupos (ver Capítulo 16, Sección 16.3).

Para algunas preguntas los estudios no aleatorizados pueden representar la mejor evidencia (de efectividad) disponible. La revisión de evidencia no aleatoria puede proporcionar una estimación de la naturaleza, la dirección y el tamaño de los efectos. La demostración de los patrones de evidencia obtenidos de los diferentes diseños de estudios puede llevar al desarrollo de diseños de estudios posteriores (incluidos ensayos aleatorizados) para probar la intervención. Los estudios generan datos cualitativos que también pueden ser relevantes para otros tipos de preguntas más allá de las preguntas de efectividad. Por ejemplo, se pueden obtener datos sobre las preferencias de quienes probablemente reciban las intervenciones y los factores que limitan o facilitan el resultado exitoso de intervenciones específicas. Hay investigaciones en curso acerca de las diferencias entre los estudios aleatorizados y no aleatorizados de intervenciones en salud pública y promoción de salud (por ejemplo, el *Methodology Programme* del RU). El capítulo 13 discute aspectos generales sobre la inclusión de estudios no aleatorizados en las revisiones Cochrane y el Capítulo 20 aborda los estudios cualitativos.

21.3 Búsqueda

Encontrar estudios sobre intervenciones en salud pública y promoción de salud es mucho más complicado que recuperar estudios médicos debido a que la bibliografía es muy escasa (Peersman 2001). La naturaleza multidisciplinaria de la salud pública y la promoción de salud significa que se pueden encontrar estudios en un número de áreas diferentes y a través de un amplio rango de bases de datos electrónicas (Beahler 2000, Grayson 2003). También surgen dificultades porque la terminología es poco precisa y cambia constantemente (Grayson 2003). Por lo tanto, la búsqueda de bibliografía en salud pública y promoción de salud puede ser una tarea muy compleja y requiere que los revisores utilicen otros métodos de recuperación además de la búsqueda en las bases de datos para recuperar estudios.

Para superar algunas de las dificultades en la identificación de investigaciones cualitativas, la mejor práctica actual requiere que el investigador realice búsquedas exhaustivas (p.ej. búsquedas sensibles de fuentes múltiples). Sin embargo, este enfoque que intenta maximizar el número de registros relevantes identificados provoca la recuperación de un gran número de registros, muchos de los cuales no son relevantes (Shaw 2004). Debido a los términos de indexación inadecuados para las investigaciones cualitativas en las bases de datos bibliográficas, actualmente no se recomienda que se apliquen filtros de diseños de estudio. Se reconoce que con frecuencia se necesita tomar decisiones pragmáticas cuando se equilibra el tiempo y los otros recursos requeridos en la realización de búsquedas exhaustivas contra el cociente de estudios relevantes y no relevantes identificados. Los investigadores pueden decidir que necesitan aplicar filtros de diseños de estudio y, de ser así, necesitan documentarlo cuando describan sus estrategias de búsqueda para aclarar las limitaciones potenciales de las búsquedas. En la Tabla 21.3a se muestran algunas bases de datos electrónicas relevantes para una variedad de tópicos en salud pública y promoción de salud.

Tabla 21.3.a: Bases de datos relevantes para salud pública y promoción de salud (sitios web enumerados por bases de datos libremente disponibles vía Internet)

Campo	Recurso
Psicología	PsycINFO/PscyLIT
Biomédico	CINAHL, LILACS (*Latin American Caribbean Health Sciences Literature*, www.bireme.br/bvs/I/ibd.htm), *Web of Science*, Medline, EMBASE, CENTRAL, SCOPUS
Sociología	*Sociofile, Sociological Abstracts, Social Science Citation Index, Social Policy and Practice*
Educación	ERIC (*Educational Resources Information Center*), C2-SPECTR (*Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register*, www.campbellcollaboration.org), REEL (*Research Evidence in Education Library*, EPPI-Centre, eppi.ioe.ac.uk)
Transporte	NTIS (*National Technical Information Service*), TRIS (*Transport Research Information Service*, ntl.bts.gov/tris), IRRD (*International Road Research Documentation*), TRANSDOC (from ECMT (*European Conference of Ministers of Transport*).
Actividad Física	SportsDiscus
SP/PS	BiblioMap, TRoPHI (*Trials Register of Promoting Health Interventions*)

	and DoPHER (*Database of Promoting Health Effectiveness Reviews*) (EPPI-Centre, eppi.ioe.ac.uk), *Public Health Electronic Library* (*National Institute for Health and Clinical Excellence*, www.nice.org.uk/guidance) *Database of abstracts of reviews of effectiveness* (DARE)
Otro	Popline (*population health, family planning*) db.jhuccp.org/popinform/basic.html, Enviroline (environmental health) – available on Dialog, Toxfile (toxicology) – available on Dialog, Econlit (economics), NGC (*National Guideline Clearinghouse*, www.guideline.gov)
Cualitativo	ESRC *Qualitative Data Archival Resource Centre* (QUALIDATA, www.qualidata.essex.ac.uk), *Database of Interviews on Patient Experience* (DIPEX, www.dipex.org)

21.4 Evaluación de la calidad del estudio y el riesgo de sesgo

La evaluación de la calidad de los estudios de salud pública y promoción de salud, y su riesgo de sesgo resultante puede ser difícil, especialmente debido a la amplia variedad de diseños de estudio utilizados. Los revisores necesitan considerar los criterios que se utilizarán para evaluar la calidad en la etapa de planificación de la revisión. Los revisores se deberían guiar por el Grupo de Revisiones Cochrane que edite su revisión y por las herramientas de evaluación que utilice. No obstante, a continuación se describen las herramientas que pueden ser útiles para evaluar los estudios de intervenciones en salud pública y promoción de salud:

- El riesgo de sesgo en los ensayos aleatorizados se debería evaluar mediante la herramienta de la Colaboración "Riesgo de Sesgo" descrita en el Capítulo 8 (Sección 8.5).
- Los aspectos relacionados con los ensayos aleatorizados por grupos se analizan en el Capítulo 16 (Sección 16.3.2).
- Para el riesgo de sesgo en los estudios no aleatorizados los revisores deberían consultar el Capítulo 13 (Sección 13.5).
- Los revisores pueden elegir utilizar la Herramienta de Evaluación de la Calidad para los Estudios Cuantitativos (Proyecto Práctica de Salud Pública Efectiva 2007) (*Quality Assessment Tool for Quantitative Studies* (*Effective Public Health Practice Project 2007*)). Esta herramienta fue desarrollada por el Proyecto Práctica de Salud Pública Efectiva (*Effective Public Health Practice Project*), Canadá y abarca cualquier diseño de estudio cuantitativo. Completar la herramienta toma entre diez y 15 minutos. En el sitio web (http://www.myhamilton.ca/myhamilton/CityandGovernment/HealthandSocialServices/Research/EPHPP/) se publicó también un diccionario exhaustivo para la herramienta de evaluación. Esta herramienta incluye los componentes de integridad de la intervención y se consideró adecuada para su uso en las revisiones sistemáticas de efectividad en la revisión realizada por Deeks y cols. (Deeks 2003).
- Hay una guía disponible del Grupo Cochrane para una Práctica y Organización Sanitaria Efectivas (*Cochrane Effective Practice and Organisation of Care Group*) sobre los estudios de series de tiempo interrumpidas y los estudios controlados antes y después (Cochrane EPOC Group 2008).
- Los resultados de los estudios no controlados (también llamados estudios antes y después sin un grupo control) se deberían tratar con precaución. La falta de un grupo de comparación hace imposible conocer qué habría ocurrido sin la intervención. Algunos de los problemas específicos con la interpretación de los datos de los estudios no

controlados incluyen la susceptibilidad a los problemas con los factores de confusión (que incluyen las variaciones estacionales) y la regresión a la media.

21.5 Ética y desigualdades

Las intervenciones en salud pública y promoción de salud tienen la posibilidad de mejorar la salud de las poblaciones. Las revisiones sistemáticas pueden determinar la efectividad de estas intervenciones para lograr los resultados deseados. Existen algunas consideraciones éticas específicas que se deberían tomar en cuenta al revisar la efectividad de las intervenciones en salud pública y promoción de salud. Habitualmente la efectividad se mide en término del número total (población) que se beneficia con la intervención. Este enfoque consecuencialista no considera la distribución de los beneficios (Hawe 1995), por lo que no aborda aspectos relacionados con la equidad en salud. Las mejorías generales en las conductas de salud o en los resultados de salud actualmente pueden ocultar las diferencias en los resultados de salud entre los grupos (Macintyre 2003). Las intervenciones que funcionan para personas en posiciones socio-económicas medias y altas pueden no ser efectivas para las personas marginadas. Incluso las intervenciones bien intencionadas actualmente pueden aumentar las desigualdades. Las diferencias en salud que existen entre los grupos se pueden deber a interacciones complejas entre muchos de los factores relacionados con las desventajas (Jackson 2003).

Las revisiones sistemáticas de intervenciones en salud pública y promoción de salud tienen la posibilidad de investigar resultados diferenciales para grupos con variados niveles de desigualdad. Esto es importante ya que identificar el efecto de las intervenciones sobre los grupos con desigualdades permite diseñar estrategias con el objetivo de reducir las desigualdades y las inequidades en salud. Las desigualdades son “diferencias, variaciones y disparidades en los logros en la salud de los individuos y grupos” (Kawachi 2002). La equidad en salud es un concepto ético que se refiere a la imparcialidad o parcialidad de desigualdades específicas en salud. La Sociedad Internacional para la Equidad en Salud (*International Society for Equity in Health*) define la equidad en salud como: “la falta de diferencias sistemáticas potencialmente remediables en uno o más aspectos del estado de salud entre poblaciones o subgrupos definidos social, económica, demográfica o geográficamente” (Macinko 2002). Dicho de otra manera, las desigualdades en salud son aquellas desigualdades que son inmerecidas o injustas, o provienen de algún tipo de injusticia (Kawachi 2002). Las revisiones de efectividad de las intervenciones de salud pública y promoción de salud pueden proporcionar información acerca de los efectos de las intervenciones sobre las desigualdades en salud. Esta información se puede utilizar para abordar las inequidades en salud.

Las desventajas se pueden considerar en términos de lugar de residencia, raza o etnicidad, ocupación, género, religión, educación, posición socio-económica y capital social, lo que en inglés se conoce como el acrónimo PROGRESS (Evans 2003). Los revisores deberían considerar cuidadosamente cuáles de estos factores son relevantes para su población de interés, y extraerán los datos según éstos. El Área Cochrane de Equidad en Salud y el Grupo Campbell de Métodos de Equidad trabajan en las definiciones de equidad relevantes para las revisiones Cochrane: www.equity.cochrane.irg.au/en/index.html.

Las revisiones sistemáticas se basan en que haya suficientes detalles en los datos del estudio que permitan la identificación de subgrupos relevantes para el análisis en relación con las desigualdades en salud. Esto requiere prestar atención no sólo a los niveles de beneficio o daño, sino también a la distribución de los mismos; ¿quién se beneficia, quién se daña, quién es excluido?

Las revisiones de la efectividad de las intervenciones relacionadas con las desigualdades en salud necesitan tres componentes para el cálculo:

- una medida válida del estado de salud (o cambio en el estado de salud);
- una medida de posición socio-económica (o desventaja); y
- un método estadístico para resumir la magnitud de las diferencias en la salud entre personas de diferentes grupos.

Los revisores deberían decidir cuál/es indicador/es de desigualdad o estado son relevantes para el tema de revisión. Existen muchos factores relacionados con las desigualdades (acrónimo PROGRESS) y los revisores necesitarán obtener datos de cualquiera de los factores que probablemente sean relevantes para su población de interés (PROGRESS = residencia, raza o etnicidad, ocupación, género, religión, educación, posición socio-económica y capital social).

Realizar revisiones que aborden las desigualdades es complicado y no sólo por la obtención limitada de información acerca de las diferencias entre los grupos, sino también por el hecho de que existe participación limitada de los grupos con desventaja en la investigación. A pesar de estas barreras, las revisiones sistemáticas pueden desempeñar un papel importante en el aumento de de la conciencia de las desigualdades en salud. El Área Cochrane de Equidad en Salud y el Grupo Campbell de Métodos de Equidad han identificado varias revisiones relevantes para la equidad que pueden proporcionar una guía adicional para los revisores.

Para localizar estudios que examinen las desigualdades, los revisores necesitarán revisar ampliamente la red cuando realicen las búsquedas y establecer contacto con los autores para información adicional acerca de datos socioeconómicos. Esta última tarea puede ser necesaria porque a menudo los estudios primarios no presentan información sobre la composición socioeconómica de los participantes (Oakley 1998, Jackson 2003, Ogilvie 2004). Una vez que los estudios se han evaluado y se han extraído los datos, es necesario clasificar los estudios como si fueran efectivos para reducir las desigualdades en salud. Una intervención efectiva para reducir las desigualdades generalmente es la que es más efectiva para grupos o individuos con desventaja. Una intervención potencialmente efectiva para reducir las desigualdades es una que es igualmente efectiva en el espectro socioeconómico (puede reducir las desigualdades en salud debido a que la prevalencia de los problemas de salud entre las personas con desventaja es mayor). La valoración se hace más difícil cuando la intervención se dirige solamente a individuos o grupos con desventaja. En una revisión Cochrane de problemas de alimentación en escuelas, las intervenciones efectivas dirigidas solamente a niños con desventaja se etiquetaron como “potencialmente” efectivas para reducir la desigualdades socioeconómicas en salud (Kristjansson 2007). No es posible determinar la efectividad diferencial si los estudios incluyen niveles mixtos de ventajas y desventajas pero no incluyen resultados que se puedan separar por grupos socioeconómicos (u otro similar).

21.6 Contexto

El tipo de intervenciones implementadas y su éxito o fracaso posterior dependen en gran medida del contexto social, económico y político en el cual se desarrollan e implementan (ver ejemplo en la Figura 21.6.a). Un problema al revisar las intervenciones en salud pública y promoción de salud es cómo diferenciar los efectos de la “intervención” de los efectos a los que sería más apropiado llamar “programa por interacciones de contexto” (Hawe 2004). Tradicionalmente los resultados se han atribuido a la intervención. Sin embargo, los resultados señalados en los estudios pueden de hecho deberse a factores

preexistentes del contexto en el cual se introdujo la intervención. Por lo tanto, el contexto sólo se debería considerar y medir como un modificador del efecto en los estudios (Eccles 2003, Hawe 2004). Estos factores contextuales se pudieran relacionar con aspectos de la "organización huésped" del programa. Los aspectos más amplios del contexto pudieran incluir aspectos del *sistema dentro del cual opera la organización huésped*. Algunos investigadores también argumentarían que los factores contextuales también pertenecen a las *características del grupo o población objetivo*. Durante muchos años se han reconocido estos aspectos (pero no se han especificado claramente) cuando quienes toman las decisiones han planteado que los resultados de las revisiones de evidencia de otros países no se aplican al suyo propio.

El uso del término "evaluación contextual" se ha hecho más prevalente en la promoción de salud después de la revisión realizada por Israel y cols. (Israel 1995). Sin embargo, las investigaciones sistemáticas de las interacciones a nivel contextual como parte del diseño de los ensayos aleatorizados de intervenciones a nivel comunitario u organizacional son casi desconocidas (Eccles 2003, Hawe 2004). En su lugar, los aspectos contextuales se han explorado como parte del campo más desarrollado de la investigación en sostenibilidad o en programas de institucionalización: ver Sección 21.7. Un campo de investigación multidisciplinario relacionado y en crecimiento son las ciencias de la implementación y la integración que llevan a los investigadores más hacia la complejidad de los procesos de cambio que lo que representan las intervenciones (Ottoson 1987, Bauman 1991, Scheirer 1994). Actualmente los estudios cuantitativos han quedado por detrás de los análisis cualitativos contextuales.

Es muy difícil separar sistemáticamente los efectos contextuales de los efectos de la investigación en un estudio que no se haya diseñado para este fin. Aunque algunos programas han sido transferidos de un contexto a otro y se han observado beneficios (Resnicow 1993), en otros no (Lumley 2004). Es de esperar (en teoría) que los diseños aleatorizados por grupo equilibren importantes aspectos contextuales, siempre que el tamaño de muestra sea suficiente. Sin embargo, actualmente pocos investigadores miden o describen aspectos contextuales que pudieran ser importantes para la evaluación. También se han hecho llamadas recientes a un mayor énfasis en la validez externa (Glasgow 2006, Green 2006). En su trabajo conjunto, se estimula a los editores de revistas e investigadores a que examinen en más detalle y describan aspectos del contexto de la intervención (Armstrong 2008). Esto se debería reflejar en el contenido de las revisiones Cochrane futuras.

Figura 21.6.a: Ejemplo de éxito de una intervención según el contexto en el cual se implementó (Frommer 2003)

Intervención basada en los medios de comunicación para promover el consumo de frutas y vegetales
↓ *Según los siguientes factores contextuales:*
Disponibilidad y precio relativos de las frutas y vegetales
↓ *Según los siguientes factores contextuales:*
Factores geográficos, sistemas de distribución de alimentos y precios minoristas

21.7 Sostenibilidad

La sostenibilidad se refiere al fenómeno general de la continuación de una intervención o sus efectos (Shediac-Rizkallah 1998, Swerissen 2004). La sostenibilidad de las intervenciones debe ser una consideración importante en las revisiones sistemáticas. Es probable que la atención a la viabilidad a largo plazo de las intervenciones sanitarias aumente cuando se incremente la preocupación por asignar los escasos recursos de forma efectiva y eficiente por parte de los elaboradores de políticas sanitarias, los médicos y los patrocinadores (Shedic-Rizkallah 1998). Los usuarios de las revisiones están interesados en conocer si los beneficios para la salud, como las reducciones en las enfermedades específicas o las mejorías en la salud, se van a sustentar más allá del período de aplicación de las intervenciones.

Desafortunadamente, a menudo no se realiza la obtención de datos sobre el grado en el cual se sustentan la intervención y los resultados, lo que limita el grado en el cual es posible evaluar las repercusiones a largo plazo. Una consideración cuidadosa en las revisiones Cochrane de cómo los estudios anteriores han (o no han) abordado los aspectos de la sostenibilidad aumentará el conocimiento en esta área y es de esperar que también estimule la mejoría en el diseño para la evaluación de la sostenibilidad en estudios futuros.

Un programa sostenido o sostenible no da lugar necesariamente a resultados sostenibles y no todas las intervenciones necesitan ser sostenibles para que sean útiles o efectivas (Shediac-Rizkallah 1998). Además, los revisores deberían considerar si la sostenibilidad de los resultados es relevante para los objetivos de la intervención. Si este es el caso los revisores deberían considerar qué resultados se han (o se deberían haber) medido, en qué período y qué patrón de resultados se mantiene en el tiempo.

Se debería buscar información sobre los factores contextuales y las características de la intervención que puedan explicar el grado en el cual se sostienen las intervenciones o los resultados. Cuando no se ha medido la sostenibilidad de los resultados los revisores deberían explorar la *posibilidad* de que los resultados de las intervenciones se sostengan. Existen cuatro marcos que pueden ser útiles para determinar la sostenibilidad:

1. Bossert enumera los cinco factores siguientes que influyen en la sostenibilidad (Bossert 1990):
   - las variables económicas y políticas alrededor de la implementación y evaluación de la intervención;
   - la fortaleza de la institución que implementa la intervención;
   - la integración completa de las actividades a los programas/servicios/currículo/etc. existentes;
   - si el programa incluye un fuerte componente de entrenamiento (construcción de capacidades); y
   - la implicación/participación de la comunidad en el programa.

2. El marco desarrollado por Swerissen y Crisp (Swerissen 2004) guía las decisiones acerca de la probable sostenibilidad de las intervenciones y los efectos a diferentes niveles de organización social. Este marco esboza las relaciones entre el nivel de intervención, las estrategias y la probable sostenibilidad de las intervenciones y los efectos.

3. Shediac-Rizkallah y Bone presentan un marco útil para conceptualizar la sostenibilidad (Shediac-Rizkallah 1998). En este marco los aspectos clave de la sostenibilidad del programa se definen como 1) mantenimiento de los beneficios sanitarios del programa; 2) institucionalización de un programa dentro de una organización; y 3) construcción de

capacidades en la comunidad que recibe el programa. Los factores clave que influyen en la sostenibilidad se definen como 1) factores en el medio más amplio; 2) factores dentro del ámbito organizacional; y 3) diseño del proyecto y factores de implementación.

4. El *Centre for Health Promotion*, Universidad de Toronto, también ha producido un documento que esboza cuatro componentes integrados de la sostenibilidad (Health Communication Unit 2001).

21.8 Aplicabilidad y transferibilidad

Es necesario considerar la aplicabilidad cuando se decide cómo transferir los hallazgos de un estudio o una revisión determinados a una población, intervención o ámbito específicos (ver Capítulo 12, Sección 12.3). La *transferibilidad* o *posibilidad de transferencia* son términos similares y apropiados. La aplicabilidad está muy relacionada con la integridad, el contexto y la sostenibilidad, como se discute en las secciones anteriores de este capítulo.

Las revisiones sistemáticas de salud pública y promoción de salud incluyen varios aspectos que hacen que el proceso de determinar la aplicabilidad sea incluso más complejo que en la bibliografía de los ensayos clínicos. Primero, varias intervenciones en salud pública no incluyen la asignación al azar. Aunque no es una característica inherente de los diseños no aleatorizados, estos estudios pueden tener criterios de elegibilidad, ámbitos e intervenciones menos definidos, lo que hace que la determinación de la aplicabilidad sea más difícil. Entonces, nuevamente los resultados de los ensayos aleatorizados pueden ser menos generalizables debido a que quienes administran la intervención o los participantes de los estudios no son representativos ni típicos del grupo objetivo (Black 1996). En segundo lugar, las intervenciones en salud pública y promoción de salud tienden a tener múltiples componentes. Esto dificulta 1) determinar qué componente específico de la intervención tiene el efecto observado, y 2) evaluar la sinergia entre los componentes. En tercer lugar, en las intervenciones comunitarias la implementación y el cumplimiento pueden ser mucho más difíciles de alcanzar y medir. Esto hace más ardua la interpretación y aplicación de los hallazgos. Cuarto, en las intervenciones de salud pública y promoción de salud las características socioculturales subyacentes de las comunidades son complejas y difíciles de medir. Por lo tanto, es complicado definir a quién y en qué grado se aplicó la intervención, lo que complica la determinación de la aplicabilidad. Por otra parte, esta heterogeneidad puede incrementar la aplicabilidad, ya que las poblaciones, ámbitos e intervenciones originales pueden ser bastante diversos, lo que aumenta la probabilidad de que la evidencia se pueda aplicar ampliamente.

Los revisores están en una posición ideal para resumir los diferentes aspectos de la evidencia relevante para los posibles usuarios. Esto permite que los usuarios comparen su situación o ámbito con el que se presenta en la revisión y noten las semejanzas y diferencias. Los usuarios pueden entonces ser explícitos acerca de las relaciones entre el conjunto de evidencia y su situación específica.

Las siguientes preguntas pueden ayudar a los revisores a considerar aspectos de la aplicabilidad y transferibilidad relevantes para la salud pública y la promoción de salud (Wang 2006).

Aplicabilidad

- ¿El medio político de la sociedad local permite implementar esta intervención?
- ¿Existe alguna barrera política para implementar esta intervención?

- ¿El público general y la (sub) población objetivo aceptaría esta intervención? ¿Algún aspecto de la intervención se opone a las normas sociales locales? ¿Es éticamente aceptable?
- ¿Los contenidos de la intervención se pueden modificar para ajustarlos a la cultura local?
- ¿Los recursos fundamentales para implementar esta intervención están disponibles en el ámbito local? (una lista de los recursos fundamentales puede ayudar a responder esta pregunta);
- ¿La población objetivo en el ámbito local tiene un nivel educacional suficiente para comprender los contenidos de la intervención?
- ¿Qué organización será responsable de administrar esta intervención en el ámbito local?
- ¿Existe alguna posible barrera para implementar esta intervención debido a la estructura de la organización?
- ¿Los proveedores de la intervención en el ámbito local tienen las habilidades para administrar esta intervención? De no ser así, ¿habrá una formación disponible?

Transferabilidad

- ¿Cuál es la prevalencia inicial del problema de salud de interés en el ámbito local? ¿Cuál es la diferencia en la prevalencia entre el ámbito del estudio y el ámbito local?
- ¿Las características de la población objetivo son comparables entre el ámbito del estudio y el ámbito local? Con respecto a los aspectos específicos que se abordarán en la intervención, ¿es posible que las características de la población objetivo, como la etnicidad, el estado socioeconómico, el nivel educacional, etc. repercutan sobre la efectividad de la intervención?
- ¿La capacidad para implementar la intervención es comparable entre el ámbito del estudio en aspectos como el medio político, la aceptabilidad social, los recursos, la estructura organizacional y las habilidades de los proveedores locales?

21.9 Información del capítulo

Editores: Rebecca Armstrong, Elizabeth Waters y Jodie Doyle.

La versión en inglés de este capítulo se debería citar como: Armstrong R, Waters E, Doyle J (editors). Chapter 21: Reviews in health promotion and public health. In Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Autores que contribuyeron: Rebecca Armstrong, Elizabeth Waters, Nicki Jackson, Sandy Oliver, Jennie Popay, Jonathan Shepherd, Mark Petticrew, Laurie Anderson, Ross Bailie, Ginny Brunton, Penny Hawe, Elizabeth Kristjansson, Lucio Naccarella, Susan Norris, Elizabeth Pienaar, Helen Roberts, Wendy Rogers, Amanda Sowden y Helen Thomas.

21.10 Referencias

Armstrong 2008
Armstrong R, Waters E, Moore L, Riggs E, Cuervo LG, Lumbiganon P, Hawe P. Improving the reporting of public health intervention research: advancing TREND and CONSORT. *Journal of Public Health (Oxford)* (in press, 2008).

Bauman 1991
Bauman LJ, Stein RE, Ireys HT. Reinventing fidelity: the transfer of social technology among settings. *American Journal of Community Psychology* 1991; 19: 619-639.

Beahler 2000
Beahler CC, Sundheim JJ, Trapp NI. Information retrieval in systematic reviews: challenges in the public health arena. *American Journal of Preventive Medicine* 2000; 18: 6-10.

Black 1996
Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; 312: 1215-1218.

Bossert 1990
Bossert TJ. Can they get along without us? Sustainability of donor-supported health projects in Central America and Africa. *Social Science and Medicine* 1990; 30: 1015-1023.

Cochrane EPOC Group 2008
Cochrane EPOC Group. Cochrane Effective Practice and Organisation of Care Group. Available from: http://www.epoc.cochrane.org (accessed 1 January 2008).

Deeks 2003
Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG. Evaluating non-randomised intervention studies. *Health Technology Assessment* 2003; 7: 27.

Donner 2004
Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *American Journal of Public Health* 2004; 94: 416-422.

Eccles 2003
Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Quality and Safety in Health Care* 2003; 12: 47-52.

Effective Public Health Practice Project 2007
Effective Public Health Practice Project. Effective Public Health Practice Project [Updated 25 October 2007]. Available from: http://www.city.hamilton.on.ca/PHCS/EPHPP (accessed 1 January 2008).

Evans 2003
Evans T, Brown H. Road traffic crashes: operationalizing equity in the context of health sector reform. *Injury Control and Safety Promotion* 2003; 10: 11-12.

Frommer 2003
Frommer M, Rychetnik L. From evidence-based medicine to evidence-based public health.

In: Lin V, Gibson B (editors). *Evidence-based Health Policy: Problems and Possibilities*. Melbourne (Australia): Oxford University Press, 2003.

Glasgow 2006
Glasgow RE, Green LW, Klesges LM, Abrams DB, Fisher EB, Goldstein MG, Hayman LL, Ockene JK, Orleans CT. External validity: we need to do more. *Annals of Behavioral Medicine* 2006; 31: 105-108.

Glasziou 2004
Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004; 328: 39-41.

Grayson 2003
Grayson L, Gomersall A. *A Difficult Business: Finding the Evidence for Social Science Reviews*. London (UK): ESRC UK Centre for Evidence Based Policy and Practice, 2003.

Green 2006
Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. *Evaluation and the Health Professions* 2006; 29: 126-153.

Hawe 1995
Hawe P, Shiell A. Preserving innovation under increasing accountability pressures: the health promotion investment portfolio approach. *Health Promotion Journal of Australia* 1995; 5: 4-9.

Hawe 2004
Hawe P, Shiell A, Riley T, Gold L. Methods for exploring implementation variation and local context within a cluster randomised community intervention trial. *Journal of Epidemiology and Community Health* 2004; 58: 788-793.

Health Communication Unit 2001
Health Communication Unit. Overview of Sustainability [Version 8.2, 30 April 2001]. Available from: http://www.thcu.ca/infoandresources/sustainability.htm (accessed 1 January 2008).

Israel 1995
Israel BA, Cummings KM, Dignan MB, Heaney CA, Perales DP, Simons-Morton BG, Zimmerman MA. Evaluation of health education programs: current assessment and future directions. *Health Education Quarterly* 1995; 22: 364-389.

Jackson 2003
Jackson T, Aldrich R, Dixon J, Furler J, Turrell G, Wilson A, Duell N, Robertson L, Leonard J. *Using Socioeconomic Evidence in Clinical Practice Guidelines*. Canberra (Australia): National Health and Medical Research Council, 2003.

Kawachi 2002
Kawachi I, Subramanian SV, Almeida-Filho N. A glossary for health inequalities. *Journal of Epidemiology and Community Health* 2002; 56: 647-652.

Kristjansson 2007
Kristjansson EA, Robinson V, Petticrew M, MacDonald B, Krasevec J, Janzen L, Greenhalgh T, Wells G, MacGowan J, Farmer A, Shea BJ, Mayhew A, Tugwell P. School feeding for improving the physical and psychosocial health of disadvantaged elementary school children. *Cochrane Database of Systematic Reviews* 2007, Issue 1. Art No: CD004676.

Lumley 2004
Lumley J, Oliver SS, Chamberlain C, Oakley L. Interventions for promoting smoking cessation during pregnancy. *Cochrane Database of Systematic Reviews* 2004, Issue 4. Art No: CD001055.

Macinko 2002
Macinko JA, Starfield B. Annotated Bibliography on Equity in Health, 1980-2001. *International Journal for Equity in Health* 2002; 1: 1.

Macintyre 2003
Macintyre S. Evaluating the evidence on measures to reduce inequalities in health. In: Oliver A, Exworthy M (editors). *Health Inequalities: Evidence, Policy and Implementation. Proceedings from a meeting of the Health Equity Network*. London (UK): The Nuffield Trust, 2003.

Oakley 1998
Oakley A, Peersman G, Oliver S. Social characteristics of participants in health promotion effectiveness research; trial and error? *Education for Health* 1998; 11: 305-317.

Ogilvie 2004
Ogilvie D, Petticrew M. Reducing social inequalities in smoking: can evidence inform policy? A pilot study. *Tobacco Control* 2004; 13: 129-131.

Ottoson 1987
Ottoson JM, Green LW. Reconciling concept and context: theory of implementation. In: Ward WB (editors). *Advances in Health Education and Promotion Volume 2*. Greenwich (CT): JAI Press, 1987.

Peersman 2001
Peersman G, Oakley A. Learning from research. In: Oliver S, Peersman G (editors). *Using Research for Effective Health Promotion*. Buckingham (UK): Open University Press, 2001.

Petticrew 2003
Petticrew M, Roberts H. Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology and Community Health* 2003; 57: 527-529.

Resnicow 1993
Resnicow K, Cross D, Wynder E. The Know Your Body program: a review of evaluation studies. *Bulletin of the New York Academy of Medicine* 1993; 70: 188-207.

Scheirer 1994
Scheirer MA. Designing and using process evaluations. In: Wholey JS, Hatry HP, Newcomer KE (editors). *Handbook of Practical Program Evaluation*. San Francisco: Jossey Bass, 1994.

Shaw 2004
Shaw RL, Booth A, Sutton AJ, Miller T, Smith JA, Young B, Jones DR, xon-Woods M. Finding qualitative research: an evaluation of search strategies. *BMC Medical Research Methodology* 2004; 4: 5.

Shediac-Rizkallah 1998
Shediac-Rizkallah MC, Bone LR. Planning for the sustainability of community-based health programs: conceptual frameworks and future directions for research, practice and policy. *Health Education Research* 1998; 13: 87-108.

Swerissen 2004
Swerissen H, Crisp BR. The sustainability of health promotion interventions for different levels of social organization. *Health Promotion International* 2004; 19: 123-130.

Wang 2006
Wang S, Moss JR, Hiller JE. Applicability and transferability of interventions in evidence-based public health. *Health Promotion International* 2006; 21: 76-83.

ÍNDICE
CAPÍTULO 22: RESÚMENES DE REVISIONES

PUNTOS CLAVE

22.1 INTRODUCCIÓN

22.1.1 Definición de Resumen de revisiones Cochrane
22.1.2 Justificación para los Resúmenes Cochrane
Tabla 22.1.a: Motivos para resumir las revisiones y su factibilidad para la publicación como un Resumen Cochrane

22.2 PREPARACIÓN DE UN RESUMEN DE REVISIONES COCHRANE

22.2.1 Aspectos organizativos
22.2.2 Aspectos metodológicos
Tabla 22.2.a: Comparación de los métodos de las revisiones Cochrane de intervenciones y los Resúmenes de revisiones Cochrane
22.2.3 Actualización de los Resúmenes Cochrane

22.3 FORMATO DE UN RESUMEN COCHRANE

22.3.1 Título e información de la revisión (o información del protocolo)
22.3.2 Resumen
22.3.3 Resumen en lenguaje sencillo
22.3.4 Texto de un Resumen Cochrane
22.3.5 Revisiones y referencias
22.3.5.1 Referencias a las revisiones
22.3.5.2 Otras referencias
22.3.6 Tablas
22.3.6.1 Tabla “Características de las revisiones incluidas”
Figura 22.3.a: Plantilla para una tabla “Características de las revisiones incluidas”
22.3.6.2 Tabla “Resumen de las revisiones”
Figura 22.3.b: Plantilla de una tabla “Resumen de revisiones”
22.3.6.3 Otras tablas
22.3.7 Figuras
Figura 22.3.c: Ejemplo de “diagrama de bosque superior” que compara intervenciones para la enuresis en niños. Este ejemplo se preparó mediante Microsoft Excel

22.4 INFORMACIÓN DEL CAPÍTULO

22.5 REFERENCIAS

CAPÍTULO 22: RESÚMENES DE REVISIONES

Autores: Lorne A Becker y Andrew D Oxman.

Puntos clave

- Los Resúmenes de revisiones Cochrane (Resúmenes) tienen como objetivo principal resumir múltiples revisiones Cochrane de intervenciones y abordar el efecto de dos o más intervenciones potenciales sobre un problema de salud o afección únicos.
- A falta de una revisión Cochrane de intervenciones que sea relevante, los Resúmenes Cochrane pueden incluir de forma adicional revisiones sistemáticas publicadas en otros lugares.
- Los Resúmenes se deberían realizar en áreas prioritarias en las que existan varias revisiones Cochrane de intervenciones.
- Los Resúmenes tienen una estructura similar a las revisiones de intervención, pero incluyen revisiones en lugar de estudios primarios.
- Los Resúmenes incluyen una tabla “Resumen de revisiones” diseñada para reflejar las tablas “Resumen de los hallazgos” en las revisiones Cochrane de intervenciones.
- Los Resúmenes se deberían actualizar cuando se actualicen las revisiones incluidas.

22.1 Introducción

22.1.1 Definición de Resumen de revisiones Cochrane

Los Resúmenes de revisiones Cochrane (Resúmenes Cochrane) son revisiones Cochrane diseñadas para compilar la evidencia de múltiples revisiones sistemáticas de intervenciones en un documento accesible y utilizable. Este capítulo esboza la justificación para los Resúmenes Cochrane y detalla los métodos que los revisores y los Grupos Cochrane de Revisión (GCR) deberían seguir para completar estas revisiones.

22.1.2 Justificación para los Resúmenes Cochrane

Los Resúmenes Cochrane tienen como objetivo principal resumir múltiples revisiones Cochrane de intervenciones y abordar el efecto de dos o más intervenciones potenciales sobre un problema de salud o afección únicos. Los Resúmenes Cochrane destacan las revisiones Cochrane que abordan estas intervenciones potenciales y resumen sus resultados importantes.

Es importante señalar que hay otros motivos para realizar resúmenes de revisiones. Los Resúmenes de revisiones Cochrane pueden cumplir algunos, pero no todos, estos objetivos. La Tabla 22.1.a esboza los diferentes motivos para resumir las revisiones sistemáticas e indica cuáles de ellas son adecuadas para su publicación como un Resumen Cochrane. Antes de registrar o publicar un Resumen Cochrane, los GCR deberían asegurar que un Resumen planificado es adecuado para publicación.

Como se puede deducir de la Tabla 22.1.a, un objetivo central de los Resúmenes Cochrane es servir como un “primer recurso accesible” con *The Cochrane Library* y permitirle al lector un vistazo general rápido (y una lista exhaustiva) de las revisiones Cochrane pertinentes para una decisión específica. Los auditorios primarios planificados son las personas que toman decisiones (como los médicos, los elaboradores de políticas y los consumidores informados) que acceden a *The Cochrane Library* en busca de evidencia sobre un problema específico. Una vez completados, los Resúmenes Cochrane se publicarán como parte de la *Base de Datos Cochrane de Revisiones Sistemáticas* (*Cochrane Database of Systematic Reviews*) en un formato que les permita a los lectores distinguirlos fácilmente de las revisiones Cochrane de intervenciones, las revisiones de las pruebas diagnósticas y de las revisiones de metodología.

Tabla 22.1.a: Motivos para resumir las revisiones y su factibilidad para la publicación como un Resumen Cochrane

Objetivo	Criterios de selección	Ejemplos de resúmenes	Adecuado para inclusión como un Resumen de revisiones Cochrane	Comentarios
Resumir la evidencia de más de una revisión sistemática de intervenciones diferentes para la misma afección o problema.	Revisiones Cohrane de intervenciones.	Un Resumen Cochrane de intervenciones para la enuresis nocturna (Russell 2006)	Sí.	Este es el objetivo principal de los Resúmenes Cochrane (y debería ser referido como un Resumen de revisiones Cochrane en la sección de objetivos del resumen y del texto).
	Revisiones Cochrane de intervenciones y revisiones sistemáticas no Cochrane.	Algunos capítulos de *BMJ Clinical Evidence* y un número creciente de informes de evaluación de tecnología sanitaria (HTA, por sus siglas en inglés).	Posiblemente.	En ocasiones puede ser apropiado incluir revisiones sistemáticas no Cochrane y revisiones Cochrane, por ejemplo, si hay intervenciones importantes para las cuales se han publicado revisiones sistemáticas de buena calidad y no hay una revisión Cochrane disponible. Sin embargo, se estimula a los GCR a centrarse principalmente en Resúmenes de revisiones Cochrane porque: ▪ la búsqueda e inclusión de revisiones no Cochrane en los Resúmenes trae consigo esfuerzos y retos adicionales ▪ es posible que las revisiones no Cochrane no estén accesibles a los usuarios de The Cochrane Library ▪ el objetivo primario de los Resúmenes Cochrane es resumir las revisiones Cochrane y proporcionar un recurso accesible para el usuario.

Objetivo	Criterios de selección	Ejemplos de resúmenes	Adecuado para inclusión como un Resumen de revisiones Cochrane	Comentarios
Resumir la evidencia de más de una revisión sistemática de la misma intervención para la misma afección o problema en la que se abordan resultados diferentes en diferentes revisiones sistemáticas.	Revisiones Cochrane de intervenciones.	Un Resumen de revisiones Cochrane sobre terapia de reemplazo hormonal (TRH) para la menopausia en la que los resultados pueden incluir densidad ósea, síntomas menopáusicos, riesgo/eventos cardiovasculares, función cognitiva, etc.	Ocasionalmente.	Como regla, las revisiones Cochrane individuales deberían incluir todos los resultados que sean importantes para las personas que toman decisiones acerca de una intervención. Sin embargo, en ocasiones, como es el caso de la TRH, diferentes revisiones sistemáticas han considerado una gran cantidad de resultados diferentes.
	Revisiones Cochrane de intervenciones y revisiones no Cochrane.	Algunos capítulos de *BMJ Clinical Evidence* y algunos informes de HTA.	En pocas ocasiones.	Las consideraciones para incluir revisiones sistemáticas no Cochrane son las mismas que las señaladas anteriormente.

Objetivo	Criterios de selección	Ejemplos de resúmenes	Adecuado para inclusión como un Resumen de revisiones Cochrane	Comentarios
Resumir la evidencia de más de una revisión sistemática de la misma intervención para diferentes afecciones, problemas o poblaciones.	Revisiones Cochrane de intervenciones.	Un Resumen de revisiones Cochrane de vitamina A para diferentes poblaciones y afecciones.	Ocasionalmente.	La misma intervención o similares se pueden utilizar en ocasiones para diferentes afecciones o estudios y las revisiones se pueden centrar en diferentes poblaciones. Aunque es poco probable que un resumen de estas revisiones sea de interés para los médicos y los pacientes que deciden la mejor manera de abordar un problema específico, un resumen puede ser pertinente para los elaboradores de políticas o para abordar preguntas que se respondan en diferentes revisiones.
	Revisiones Cochrane de intervenciones y revisiones no Cochrane.		En pocas ocasiones.	Las consideraciones para incluir revisiones sistemáticas no Cochrane son las mismas que las señaladas anteriormente.

Objetivo	Criterios de selección	Ejemplos de resúmenes	Adecuado para inclusión como un Resumen de revisiones Cochrane	Comentarios
Resumir la evidencia acerca de los efectos adversos a partir de una intervención de más de una revisión sistemática del uso de la intervención para una o más afecciones.	Revisiones Cochrane de intervenciones solamente o revisiones Cochrane de intervenciones y revisiones sistemáticas no Cochrane.	Un resumen de los efectos adversos de los AINE cuando se utilizan para la osteoartritis o la artritis reumatoide o la menorragia.	En pocas ocasiones.	Aunque muchas revisiones Cochrane proporcionan información sobre los efectos adversos, pocas, si es que hay alguna, se han diseñado principalmente para evaluar las tasas de efectos adversos. Muchos efectos adversos importantes ocurren con tan poca frecuencia que no es posible evaluar con exactitud su verdadera prevalencia a partir de los resultados de ensayos controlados. Por estos motivos puede que un resumen basado solamente en revisiones sistemáticas Cochrane u otras revisiones sistemáticas de ensayos controlados no proporcione un cuadro exacto del perfil de efectos adversos de una intervención específica (a menos que las revisiones sistemáticas que se resumen se hayan diseñado específicamente para abordar las tasas de efectos adversos) (ver Capítulo 14 para información adicional sobre la descripción de los efectos adversos en las revisiones Cochrane).

Objetivo	Criterios de selección	Ejemplos de resúmenes	Adecuado para inclusión como un Resumen de revisiones Cochrane	Comentarios
Proporcionar un resumen exhaustivo de un área, con la inclusión de estudios que no se han incluido en revisiones sistemáticas.	Revisiones sistemáticas y estudios no incluidos en revisiones sistemáticas.	Algunos capítulos de *BMJ Clinical Evidence* y un número creciente de informes de HTA o un artículo sinóptico de revisión para una revista.	No.	La inclusión de estudios que no se han incluido previamente en una revisión sistemática puede ser apropiada en varias circunstancias, por ejemplo, cuando se realiza un informe de HTA, se desarrolla una guía de práctica clínica o para recursos como *BMJ Clinical Evidence*. Sin embargo, este tema está más allá del alcance de lo que se debería hacer en una Resumen Cochrane. Los autores de Resúmenes Cochrane deberían declarar cuando las revisiones incluidas no están actualizadas, particularmente si se han publicado estudios nuevos relevantes, y si existen intervenciones pertinentes para las cuales no se han publicado aún revisiones sistemáticas. Sin embargo, no deberían realizar una actualización de una revisión sistemática o una nueva revisión sistemática dentro del Resumen.

22.2 Preparación de un Resumen de revisiones Cochrane

22.2.1 Aspectos organizativos

El estímulo para comenzar un Resumen Cochrane debería ser un área prioritaria en la que existen varias revisiones Cochrane. La identificación de la necesidad de un Resumen puede provenir de un equipo de autores interesados, un GCR o una agrupación de GCR. Las áreas o Centros también pudieran establecer temas prioritarios para los Resúmenes Cochrane e intentar encontrar autores que los realicen. Los autores de revisiones Cochrane de intervenciones pueden asumir la función de autor de un Resumen si así lo desean, pero no es automáticamente necesario que lo sean. Los autores de Resúmenes deberían estar familiarizados con la metodología de las revisiones Cochrane de intervenciones, y sería ideal que hayan sido coautores de una.

Un GCR tendrá el control editorial sobre cada Resumen de revisiones; los títulos y protocolos deberían ser enviados de la misma manera que para las revisiones de intervenciones. En la mayoría de los casos es de esperar que todas las revisiones Cochrane a incluir en un Resumen provengan de un único GCR, y dicho GCR tendría la responsabilidad editorial. Si se prevé que se incluirán revisiones Cochrane de más de un GCR, por ejemplo, en Resúmenes de revisiones que aborden una intervención utilizada en el tratamiento de varias afecciones, el proceso editorial se discutirá entre los GCR pertinentes, y se tomará una decisión acerca de cuál/es GCR tendrá/n la función editorial, como ocurre actualmente con algunas revisiones cuando está involucrado más de un GCR.

Los autores de un Resumen que identifiquen estudios no incluidos en las revisiones Cochrane de intervenciones existentes pueden considerar acercarse al GCR pertinente para planificar una nueva revisión Cochrane con un alcance más amplio, actualizar una revisión Cochrane ya existente o realizar una nueva revisión Cochrane para una intervención que no se haya incluido en una revisión existente.

22.2.2 Aspectos metodológicos

Los Resúmenes Cochrane utilizan métodos diferentes a los de las revisiones Cochrane de intervenciones: sintetizan las revisiones de intervenciones existentes en lugar de encontrar y resumir o sintetizar estudios originales. En la Tabla 22.2.a se resumen las diferencias clave en los métodos de las revisiones Cochrane de intervenciones y de los Resúmenes Cochrane.

Los Resúmenes de revisiones Cochrane no tienen como objetivo repetir las búsquedas, la evaluación de la elegibilidad, la evaluación del riesgo de sesgo ni los metanálisis de las revisiones de intervenciones incluidas. Además, generalmente no tienen como objetivo identificar sistemáticamente algún estudio adicional u obtener resultados adicionales de los estudios. Los Resúmenes incluyen la evaluación de las limitaciones de las revisiones sistemáticas incluidas y pueden incluir metanálisis de las revisiones para proporcionar comparaciones indirectas de los efectos de diferentes intervenciones sobre un resultado determinado. Esto no implica que los resúmenes de revisiones sistemáticas que realicen un análisis más detallado incluida la evaluación crítica, nuevas búsquedas y nuevos análisis no sean apropiados, pero no es lo que se espera de los Resúmenes Cochrane.

Tabla 22.2.a: Comparación de los métodos de las revisiones Cochrane de intervenciones y los Resúmenes de revisiones Cochrane

	Revisiones Cochrane de intervenciones	Resúmenes de revisiones Cochrane	Comentarios con respecto a los Resúmenes de revisiones Cochrane
Objetivos.	Resumir la evidencia de estudios de los efectos de las intervenciones.	Resumir la evidencia de revisiones sistemáticas de los efectos de las intervenciones.	Apropiados cuando hay dos o más intervenciones para la misma afección o problema presentadas en diferentes revisiones Cochrane de intervenciones.
Criterios de selección.	Describir los criterios de inclusión y exclusión para los estudios.	Describir los criterios de inclusión y exclusión para las revisiones.	Se incluyen principalmente sólo revisiones Cochrane de intervenciones. En ocasiones se pueden incluir revisiones Cochrane de intervenciones y otras revisiones encontradas en *The Cochrane Library* (*Database of Abstracts of Reviews of Effects* o *Health Technology Assessment Database*). En pocas ocasiones se pueden incluir otras revisiones sistemáticas.
Búsqueda.	Búsqueda exhaustiva de estudios relevantes.	Generalmente búsqueda sólo de revisiones Cochrane de intervenciones relevantes.	Ocasionalmente se pueden buscar revisiones sistemáticas no Cochrane.
Obtención de los datos.	A partir de los estudios incluidos.	A partir de las revisiones sistemáticas incluidas.	De ser necesario, los autores de Resúmenes pueden buscar información adicional de los autores de las revisiones sistemáticas incluidas u ocasionalmente de los estudios primarios incluidos en las revisiones sistemáticas.
Evaluación de las limitaciones.	Para los estudios incluidos, es decir, riesgo de sesgo.	Para las revisiones sistemáticas incluidas.	Los autores de los Resúmenes Cochrane deberían evaluar críticamente las revisiones incluidas mediante criterios explícitos. Se deberían considerar las limitaciones generales (p.ej. si la revisión está actualizada) y las limitaciones específicas (p.ej. si una revisión sistemática tiene limitaciones relacionadas con los objetivos específicos del Resumen).

	Revisiones Cochrane de intervenciones	Resúmenes de revisiones Cochrane	Comentarios con respecto a los Resúmenes de revisiones Cochrane
Calidad de la evidencia.	Entre los estudios para cada resultado importante.	Hasta donde sea posible se debería basar en las evaluaciones presentadas en las revisiones sistemáticas incluidas.	Se recomienda que cada Resumen incluya una evaluación de la calidad de la evidencia para cada resultado importante. Si no se realizó dicha evaluación en las revisiones sistemáticas incluidas, los autores del Resumen deberían tratar de realizarla. Si la misma ya se realizó en las revisiones sistemáticas incluidas, los autores de los Resúmenes deberían evaluar críticamente las valoraciones que se hicieron y tratar de asegurar que estas valoraciones se hicieron de forma consistente entre las revisiones incluidas.
Análisis.	Síntesis de resultados de los estudios incluidos para cada resultado importante.	Resumen de los resultados de la revisión; se pueden realizar análisis adicionales para las comparaciones entre las revisiones, generalmente comparaciones indirectas de múltiples intervenciones.	Hasta donde sea posible, los autores de los Resúmenes Cochrane deberían confiar en los análisis presentados en las revisiones incluidas. En ocasiones puede ser necesario reanalizar los datos, por ejemplo, si se analizan diferentes poblaciones o subgrupos en diferentes revisiones y es posible realizar análisis comparables entre las revisiones.

22.2.3 Actualización de los Resúmenes Cochrane

La actualización regular de un Resumen Cochrane es muy importante y sigue el proceso habitual de actualización de las revisiones Cochrane (ver Capítulo 3). Un Resumen Cochrane necesitará actualizarse en cualquier momento en el que se actualicen las revisiones incluidas. En muchos casos sólo se necesitarán cambios menores al Resumen Cochrane. Por ejemplo, si no se encuentran nuevos estudios en la actualización de una revisión Cochrane de intervenciones, sólo se deberá cambiar en el Resumen la información sobre la fecha de la última actualización para dicha revisión. Sin embargo, cuando una actualización dé lugar a cambios en los resultados y las conclusiones de una revisión Cochrane incluida, serán necesarias revisiones más amplias del Resumen.

22.3 Formato de un Resumen Cochrane

22.3.1 Título e información de la revisión (o información del protocolo)

El título de un Resumen debería tener el formato: [Intervenciones o comparaciones] para [problema de salud] en O para [tipo de personas, enfermedad o problema y ámbito si se especifica].

La parte del título "Intervenciones o comparaciones" puede tomar varios formatos según el alcance de la revisión. Si se van a considerar todas las intervenciones potenciales con la evidencia de las revisiones sistemáticas, esta sección debería señalar solamente "Intervenciones para". Si el Resumen se va a limitar a un subgrupo de intervenciones potenciales, el título debería indicar el subgrupo, por ejemplo, "Intervenciones quirúrgicas para". Si se van a comparar dos tipos de intervenciones, el comparador se debería incluir en el título, por ejemplo "Intervenciones quirúrgicas o farmacológicas para".

Cualquier otra información de la revisión es la misma que para las revisiones de intervenciones, como se describe en el Capítulo 4 (Sección 4.2).

22.3.2 Resumen

El contenido de cada encabezado del resumen debería ser el siguiente:

Antecedentes: deberían ser una o dos oraciones para explicar el contexto o elaborar los fines y la justificación del Resumen.

Objetivos: debería ser un constatación explícita del objetivo primario del Resumen, de manera ideal en una oración única. De ser posible el estilo debería estar en el formato "Resumir las revisiones Cochrane que evalúen el efecto de *[Intervenciones o comparaciones]* para *[problema de salud]* para/en *[tipo de personas, enfermedad o problema y ámbito si se especifica]*".

Métodos: esta sección debería abordar de forma sucinta la estrategia de búsqueda utilizada para identificar las revisiones sistemáticas para su inclusión en el Resumen y los métodos utilizados para la obtención y el análisis de los datos. Este último aspecto se debería limitar a la descripción de las guías utilizadas para extraer los datos y evaluar la calidad y la validez de los mismos, y no incluir detalles de qué datos se extrajeron. Se debería señalar el método mediante el cual se aplicaron las guías (por ejemplo, extracción independiente por múltiples revisores).

Resultados principales: esta sección debería comenzar con el número total de revisiones sistemáticas incluidas en el Resumen, y breves detalles relevantes para la interpretación de los resultados (por ejemplo, la calidad de las revisiones sistemáticas incluidas o un comentario sobre la comparabilidad de las revisiones, de ser apropiado). Debería abordar el objetivo primario y limitarse a los principales resultados cualitativos y cuantitativos (que generalmente incluyen no más de siete resultados clave). Los resultados incluidos se deberían seleccionar según su valor esperado para ayudar a alguna persona a tomar decisiones acerca de utilizar o no una intervención específica. De ser relevante se debería señalar el número de estudios y participantes que contribuyen a los resultados separados, junto con la calidad de la evidencia específica para estos resultados. Los resultados se deberían expresar de forma narrativa y cuantitativa si los resultados numéricos no están claros o son intuitivos (como los provenientes de los análisis de diferencias de medias estandarizadas). Las estadísticas resumen en el

resumen deberían ser las mismas que las señaladas en el texto del Resumen, y se deberían presentar de una forma estándar, como "cociente de riesgos 2,31 (intervalo de confianza del 95%: 1,13 a 3,45)". Se deberían describir los efectos relativos y absolutos, de ser posible. Sin embargo, los revisores deberían ser cuidadosos al detallar los efectos absolutos cuando el riesgo del grupo control para un resultado varíe entre los estudios o las revisiones (ver Capítulo 11, Sección 11.5.5). Si los resultados generales no se calcularon en una revisión incluida, se debería proporcionar una evaluación cualitativa o una descripción del rango y el patrón de los resultados. Sin embargo, se deberían evitar "los recuentos de votos" en los cuales se describan los números de estudios (o revisiones) "positivos" y "negativos".

Conclusiones de los autores: el objetivo primario del Resumen debería ser presentar la información, en lugar de ofrecer consejos. Las conclusiones de los autores deberían ser sucintas y provenir directamente de los hallazgos del Resumen, de manera que reflejen directamente los resultados principales. Los autores deberían ser cuidadosos en no confundir una falta de evidencia con una falta de efecto. No se deberían hacer suposiciones acerca de circunstancias prácticas, valores, preferencias, equilibrios, y en general se debería evitar proporcionar consejos o recomendaciones. Se debería señalar cualquier limitación importante de los datos y los análisis. De ser relevantes, se deberían incluir conclusiones importantes acerca de implicaciones específicas para la investigación, incluidas las revisiones sistemáticas. Los autores no deberían hacer afirmaciones generales de que "se necesitan más investigaciones".

22.3.3 Resumen en lenguaje sencillo

El resumen en lenguaje sencillo (anteriormente llamado "sinopsis") tiene como objetivo sintetizar el Resumen en un estilo simple que pueda ser comprendido por los consumidores de la atención sanitaria. Ver Capítulo 4 (Sección 4.4).

22.3.4 Texto de un Resumen Cochrane

El auditorio objetivo de un Resumen Cochrane son las personas que toman decisiones acerca de la atención sanitaria (p.ej. médicos, consumidores informados y elaboradores de políticas) que ya poseen alguna comprensión básica de la enfermedad o problema subyacente y desean determinar en qué grado se han abordado en *The Cochrane Library* las intervenciones potenciales para el problema. El Resumen debería proporcionar una síntesis de los hallazgos de las revisiones Cochrane relevantes, y dirigir a los lectores a las revisiones individuales para detalles adicionales.

El texto de un Resumen Cochrane contiene varios encabezados fijos. El autor puede agregar subencabezados en cualquier punto. Algunos encabezados específicos se diseñaron como "recomendados". El contenido de las secciones recomendadas se debería incluir en todos los Resúmenes, pero el uso de subencabezados no es obligatorio y se debería evitar si hacen que las secciones individuales sean innecesariamente cortas. También se deberían proporcionar subencabezados adicionales que puedan o no ser relevantes para una revisión específica. En el resto de esta sección se señala la categoría relevante (fija, recomendada, opcional) para cada uno de los encabezados descritos.

Antecedentes [fijo, encabezado nivel 1]
Esta sección debería abordar el cuerpo de conocimiento ya formado que incluya el contexto de las revisiones Cochrane sintetizadas en el Resumen. Los antecedentes ayudan a establecer la justificación para el Resumen. La sección debería especificar la/s pregunta/s de investigación que se abordan en el Resumen, incluida una descripción clara de la afección de interés, las intervenciones, las comparaciones y los resultados considerados. Además, debería explicar por qué es importante la pregunta que se aborda. Se debería presentar de una forma que sea comprensible a los usuarios de la

atención sanitaria bajo investigación, y debería ser concisa (generalmente alrededor de una página impresa). La sección "Antecedentes" debería contener los siguientes componentes. Aunque los subencabezados no son obligatorios, se recomiendan.

Descripción de la afección [recomendado, encabezado nivel 2]
La revisión debería comenzar con una descripción breve de la afección que se aborda y su importancia. Puede incluir información acerca de la biología, el diagnóstico, el pronóstico y la importancia para la salud pública (incluida la prevalencia o la incidencia).

Descripción de las intervenciones [recomendado, encabezado nivel 2]
Esta sección debería mencionar todas las intervenciones actualmente disponibles para la afección, se hayan evaluado o no en una revisión Cochrane de intervenciones. Cuando sea razonable, el agrupamiento de las intervenciones simplificará el texto (p.ej. nombrar fármacos antiinflamatorios no esteroides en lugar de proporcionar una lista exhaustiva de dichos fármacos por nombre). Cuando sea aplicable de debería abordar la posibilidad del uso concurrente de diferentes intervenciones (p.ej. radiación más quimioterapia). Se puede mencionar el estado relativo de varias intervenciones potenciales en la práctica clínica (si es posible).

Cómo funciona la intervención [recomendado, encabezado nivel 2]
Las revisiones sistemáticas reúnen evidencia para evaluar si en realidad ocurre el efecto esperado de una intervención. Esta sección pudiera describir la justificación teórica de por qué las intervenciones bajo revisión pueden tener una repercusión sobre las personas que posiblemente reciban la atención sanitaria, por ejemplo, al relacionar una intervención farmacológica con la biología de la afección. Los autores pueden referirse al cuerpo de la evidencia empírica como las intervenciones similares que tienen repercusión o las intervenciones idénticas que tienen repercusión en otras poblaciones. Los autores también pueden citar un grupo de bibliografías que justifique la posible efectividad. Las referencias a la bibliografía existente no deberían incluir discusiones de los resultados de las revisiones sistemáticas incluidas en el Resumen ni de los estudios abordados en dichas revisiones; este material debería tratarse en la sección "Resultados".

Por qué es importante realizar este Resumen [opcional, encabezado nivel 2]
Los antecedentes ayudan a establecer la justificación para el Resumen y deberían explicar por qué son importantes las preguntas abordadas. Debería quedar claro por qué se realizó el Resumen, cuál es el auditorio objetivo y qué decisiones se propone ayudar a informar.

Objetivos [fijo, encabezado nivel 1]
Esta sección debería comenzar con un planteamiento preciso del objetivo primario de la revisión, incluida/s la/s intervención/es revisada/s y el problema objetivo. Lo anterior puede estar seguido de una serie de objetivos específicos relacionados con los diferentes grupos de participantes, las diferentes comparaciones de intervenciones o las diferentes medidas de resultado.

Métodos [fijo, encabezado nivel 1]
La sección "Métodos" en un protocolo se debería redactar en tiempo futuro. La sección "Métodos" de la revisión debería describir lo que se hizo para obtener los resultados y conclusiones de la versión actual del Resumen. No debería discutir los métodos de las revisiones sistemáticas subyacentes que se van a sintetizar. Los comentarios sobre los métodos de dichas revisiones se deberían abordar en la sección "Descripción de las revisiones incluidas". La sección "Métodos" debería estar compuesta de varias subsecciones.

Criterios para considerar las revisiones para inclusión [fijo, encabezado nivel 2]

La pregunta de investigación del Resumen debería guiar la selección de las revisiones para la inclusión, incluida una descripción clara de los participantes (afección o problema de salud), las intervenciones, los grupos de comparación y los resultados de interés. En general los Resúmenes deberían incluir todas las revisiones Cochrane que aborden una o más de las intervenciones disponibles para la afección o problema de salud que sea el tema del Resumen. Sin embargo, en algunos casos los autores del Resumen pueden desear limitar de alguna manera sus objetivos. Por ejemplo, los autores del Resumen pueden desear limitar su alcance a ciertos tipos de intervenciones (p.ej. todos los tratamientos farmacológicos, exclusión de los tratamientos no farmacológicos). Las limitaciones serían especialmente apropiadas si las revisiones Cochrane existentes abordan varias poblaciones clínicas (p.ej. grupos que difieren en cuanto a edad, raza, sexo, estadio de la enfermedad o tipos de comorbilidades). Al tomar las decisiones de agrupar o dividir, será de ayuda tener en mente la perspectiva del encargado de tomar decisiones que lee el resumen y centrarse en la información que se necesitaría para tomar una decisión individual. Por ejemplo, es probable que las revisiones Cochrane de intervenciones que abordan la prevención de una afección determinada no se deban agrupar en un Resumen único con las revisiones de intervenciones que abordan el tratamiento de la misma afección, ya que las decisiones relacionadas con la prevención y el tratamiento se toman para diferentes poblaciones. Si estas consideraciones se incluyen en la selección de las revisiones para su inclusión en el Resumen, se deberían explicar claramente en esta sección.

Si se incluyen revisiones sistemáticas no Cochrane, esta sección debería especificar los criterios que se utilizarán para determinar si las revisiones no Cochrane son revisiones sistemáticas, y los criterios que se utilizarán para determinar qué revisiones sistemáticas se incluirán cuando haya dos o más revisiones que aborden la misma pregunta.

Métodos de búsqueda para la identificación de las revisiones [fijo, encabezado nivel 2]

Esta sección debería abordar los métodos utilizados en el Resumen para encontrar revisiones Cochrane u otras revisiones sistemáticas. La búsqueda realizada será mucho más sencilla que las estrategias de búsqueda dentro de una revisión Cochrane de intervenciones, ya que la búsqueda básica para los artículos subyacentes ya se habrá realizado. Si sólo se van a incluir revisiones Cochrane en el Resumen, la búsqueda se puede realizar dentro de la *Cochrane Database of Systematic Reviews* sin la necesidad de buscar en otras bases de datos. Si se incluyen revisiones sistemáticas de otras fuentes, esta sección debería esbozar claramente las bases de datos en las que se buscó (p.ej. *Database of Abstracts of Reviews of Effects* (Petticrew 1999) y las estrategias de búsqueda y los métodos de recuperación utilizados.

Obtención y análisis de los datos [fijo, encabezado nivel 2]

Esta sección debería presentar una breve descripción de los métodos utilizados en el Resumen. Se deberían abordar los siguientes aspectos:

Selección de las revisiones [recomendado, encabezado nivel 3]

Se debería señalar el método utilizado para aplicar los criterios de selección a las revisiones identificadas en la búsqueda y si más de un revisor aplicó de forma independiente los criterios, junto con la forma en la que se resolvieron los desacuerdos.

Extracción y procesamiento de los datos [recomendado, encabezado nivel 3]

En esta sección se debería describir el método utilizado para extraer u obtener los datos de las revisiones incluidas (por ejemplo, el uso de un formulario de obtención de datos). Se debería señalar si más de un autor extrajo los datos de forma independiente, junto con la forma en la que se resolvieron los desacuerdos. De ser relevante, se deberían describir claramente los métodos para procesar los datos en preparación para el análisis. Los autores también deberían describir qué se hizo (si se hizo algo) para obtener los datos faltantes de las revisiones incluidas.

Evaluación de la calidad metodológica de las revisiones incluidas [recomendado, encabezado nivel 3]

En cada Resumen los autores deberían abordar dos evaluaciones diferentes de la calidad: la calidad metodológica de las revisiones sintetizadas en el Resumen y la calidad de la evidencia en dichas revisiones, como se describe más adelante.

Los métodos utilizados para realizar ambos tipos de evaluación se deberían describir en esta sección. Para ambas evaluaciones se recomienda que más de un revisor aplique los criterios de forma independiente. De ser así, se debería declarar, junto con la forma en la que se resolvió cualquier desacuerdo. Las herramientas utilizadas (p.ej. GRADE) se deberían describir o señalar la referencia, con la indicación de cómo se incorporan estas evaluaciones a la interpretación de los resultados del Resumen.

Calidad de las revisiones incluidas [recomendado, encabezado nivel 4]

Se deberían describir los métodos utilizados para evaluar la calidad metodológica de las revisiones incluidas en el Resumen. Se han realizado investigaciones limitadas sobre la evaluación de la calidad o del riesgo de sesgo en las revisiones sistemáticas, y no es posible recomendar un instrumento específico para valorar la calidad de las revisiones incluidas. Sin embargo, hay algunos cuestionarios y listas de verificación disponibles (Oxman 1994, Shea 2006).

Calidad de la evidencia en las revisiones incluidas [recomendado, encabezado nivel 4]

Las revisiones Cochrane de intervenciones que utilizan métodos excelentes pueden sintetizar la evidencia con limitaciones importantes, debido a los sesgos potenciales dentro y entre los estudios incluidos, resultados contradictorios entre los estudios individuales, evidencia escasa o falta de relevancia (direccionalidad) para la pregunta de revisión (ver Capítulo 12, Sección 12.2). Se deberían sintetizar los métodos utilizados en el Resumen para determinar la calidad de la evidencia en apoyo de cada conclusión del Resumen. De forma ideal, la información sobre la cual se basan estas evaluaciones debería estar disponible en las tablas “Características de los estudios incluidos”, “Riesgo de sesgo” y “Resumen de los hallazgos” proporcionadas en las revisiones incluidas. Actualmente se recomienda que las evaluaciones del riesgo de sesgo se describan de una forma estandarizada en las revisiones Cochrane (ver Capítulo 8) y que se utilice el enfoque GRADE para evaluar la calidad de la evidencia entre los estudios para cada resultado importante para las revisiones Cochrane y los Resúmenes de revisiones Cochrane (ver Capítulo 11, Sección 11.5 y Capítulo 12, Sección 12.2).

Síntesis de los datos [recomendado, encabezado nivel 3]

Muchos Resúmenes sencillamente extraerán los datos de las revisiones sistemáticas subyacentes y les darán un nuevo formato en tablas o figuras. Sin embargo, en algunos casos los Resúmenes pueden incluir comparaciones indirectas basadas en análisis estadísticos formales, especialmente si no hay evidencia de comparaciones directas (Glenny 2005). Los métodos estadísticos para realizar las comparaciones indirectas, y para los metanálisis simultáneos de intervenciones múltiples, son muy relevantes para los Resúmenes y se discuten en el Capítulo 16 (Sección 16.6). La evidencia de las comparaciones indirectas puede ser menos fiable que la evidencia de las comparaciones directas. Si ninguna de las revisiones incluidas investigó comparaciones directas, pero se conocen estudios de comparaciones directas o se cree que se han realizado, los autores de los Resúmenes no deberían intentar las comparaciones indirectas. Los autores que deseen realizar comparaciones indirectas o metanálisis de múltiples tratamientos deberían buscar apoyo estadístico y metodológico adecuado.

Cuando se utilicen más enfoques cualitativos o narrativos los revisores deberían señalar qué, métodos utilizaron, si utilizaron alguno, para estandarizar la descripción de los resultados entre las revisiones incluidas, incluida la conversión de las estadísticas

resumen y cualquier estandarización para diferentes riesgos del grupo control. Los autores deberían ser cautelosos cuando comparen los efectos absolutos entre las revisiones si existen diferencias en los riesgos del grupo control (ver Capítulo 11, Sección 11.5.5).

Resultados [fijo, encabezado nivel 1]
Descripción de las revisiones incluidas [fijo, encabezado nivel 2]
La descripción de las revisiones incluidas debería ser concisa, pero debería proporcionar detalles suficientes para permitir que el lector tenga una idea de las características de los participantes incluidos en las revisiones sintetizadas: la dosis, duración u otras características de las intervenciones. Si existen diferencias importantes entre estas revisiones incluidas (p.ej. diferencias en los criterios de la revisión para la inclusión o exclusión de los estudios, comparadores diferentes o el uso de medidas de resultado diferentes), se deberían señalar claramente. Además, se debería declarar cualquier discrepancia entre los objetivos y los criterios de elegibilidad de las revisiones incluidas y los objetivos del Resumen. Por ejemplo, los revisores pueden haber omitido los análisis de un subgrupo específico o de un resultado clave que era de particular interés para los autores del Resumen. También se debería señalar si algunas revisiones se han actualizado más recientemente que otras. Una gran parte del material de esta sección se puede resumir en la tabla “Características de las revisiones incluidas” (ver Sección 22.3.6 para más detalles).

Calidad metodológica de las revisiones incluidas [recomendado, encabezado nivel 2]
Calidad de las revisiones incluidas [recomendado, encabezado nivel 3]
La calidad general de las revisiones sistemáticas incluidas en el Resumen se puede sintetizar, incluida cualquier variabilidad entre las revisiones y cualquier deficiencia importante en las revisiones individuales. Los criterios que se utilizaron para evaluar la calidad de las revisiones se deberían describir o citar su referencia en la sección “Métodos” y no aquí. Si se considera que es importante proporcionar detalles sobre cómo se calificó cada revisión incluida con respecto a cada criterio, se debería detallar en una Tabla adicional y no describirlo en detalle en el texto.

Calidad de la evidencia en las revisiones incluidas[recomendado, encabezado nivel 3]
La calidad general de la evidencia en las revisiones incluidas se debería resumir, por ejemplo, mediante GRADE para los resultados más importantes (ver también Capítulo 13, Sección 13.2).

Efectos de las intervenciones [fijo, encabezado nivel 2]
Los hallazgos más importantes sobre los efectos de las intervenciones estudiadas se deberían resumir aquí. La sección se debería organizar alrededor de categorías clínicamente significativas en lugar de sencillamente enumerar los hallazgos de cada revisión incluida. Estas categorías pudieran incluir aspectos como los tipos de intervenciones (tratamientos farmacológicos, intervenciones quirúrgicas, intervenciones conductuales, etc.); estadios de la enfermedad (presintomática, enfermedad temprana, enfermedad avanzada); características de los participantes (edad, sexo, raza); o tipos de resultados (supervivencia, estado funcional, efectos adversos). Se estimula la utilización de subencabezados si facilitan la lectura. Los hallazgos de las revisiones individuales y cualquier resumen estadístico de las mismas se deberían incluir en las figuras o las tablas resumen.

En esta sección se debería señalar cualquier resultado que los autores del Resumen consideren importante pero para los cuales los revisores no pudieron encontrar evidencia (debido a que no se encontraron estudios o a que los estudios identificados no documentaron el resultado importante). Además, esta sección debería incluir un resumen narrativo de los resultados importantes que no es posible resumir con facilidad mediante

datos numéricos y que no es probable que se incluyan en las tablas de resultados del Resumen.

Los autores deberían evitar hacer inferencias en esta sección. Un error frecuente que se debería evitar (al describir los resultados y establecer conclusiones) es la confusión entre "no hay evidencia de un efecto" y "evidencia de que no hay efecto". Cuando no haya evidencia concluyente, es incorrecto señalar que el Resumen muestra que una intervención "no tiene efecto" o que "no es diferente" de la intervención control. En este caso es más apropiado declarar que los datos, con su intervalo de confianza, son compatibles con una reducción o un aumento en el resultado.

Discusión [fijo, encabezado nivel 1]

Resumen de los resultados principales [recomendado, encabezado nivel 2]

Proporcionar aquí una síntesis concisa de los hallazgos principales, el equilibrio entre beneficios y daños importantes y destacar cualquier incertidumbre significativa.

Compleción general y aplicabilidad de la evidencia[recomendado, encabezado nivel 2]

¿Las revisiones incluidas son suficientes para abordar todos los objetivos del Resumen? De no ser así, ¿qué lagunas existen? ¿Se investigaron todos los tipos de participantes, intervenciones y resultados relevantes? Describir la relevancia de la evidencia para la pregunta del Resumen. Esto debería dar lugar a una valoración general de la validez externa del Resumen. Aquí se pueden incluir los comentarios sobre cómo los resultados del Resumen se ajustan al contexto de la práctica actual, aunque los autores deberían tener en mente que la práctica actual puede variar internacionalmente y entre las poblaciones.

Calidad de la evidencia [recomendado, encabezado nivel 2]

¿Las revisiones incluidas en el Resumen permiten establecer conclusiones sólidas con respecto al/los objetivo/s abordado/s en el Resumen? La discusión pudiera incluir si en la revisión original se identificaron todos los estudios relevantes, si fue posible obtener todos los datos relevantes, o si los métodos utilizados (por ejemplo, la búsqueda, la selección de los estudios, la obtención y el análisis de los datos) pudieron haber introducido sesgo. Esto puede variar para diferentes intervenciones, resultados o subgrupos clínicos. De ser así, la discusión debería identificar claramente la calidad de la evidencia para cada una de las áreas clave de interés.

Sesgos potenciales en el proceso de resumen [recomendado, encabezado nivel 2]

Señalar las fortalezas y limitaciones del Resumen con respecto a evitar los sesgos. Estos pueden ser factores que están dentro o fuera del control de los autores del Resumen. La discusión pude incluir si se identificaron todas las revisiones relevantes y se incluyeron en el Resumen, si fue posible obtener todos los datos relevantes o si los métodos utilizados (por ejemplo, búsqueda, selección de los estudios, obtención y análisis de los datos) pueden haber introducido sesgo.

Acuerdo y desacuerdo con otros estudios o revisiones [recomendado, encabezado nivel 2]

Aquí se pueden incluir comentarios sobre la forma en la cual las revisiones se ajustan al contexto de otra evidencia, y señalar claramente si la otra evidencia fue sistemáticamente revisada.

Conclusiones de los autores [fijo, encabezado nivel 1]

Esta sección debería presentar las conclusiones de los autores del Resumen, no sencillamente volver a señalar las diferentes conclusiones de los autores de las revisiones incluidas/subyacentes. El objetivo primario de esta sección debería ser presentar la

información en lugar de ofrecer consejo. Las conclusiones de los autores se dividen en dos secciones, como se muestra a continuación.

Implicaciones para la práctica [fijo, encabezado nivel 2]
Las implicaciones para la práctica deberían ser tan prácticas y no ambiguas como sea posible. No deberían ir más allá de la evidencia que se revisó y deberían estar justificadas por los datos presentados en la revisión. “No hay evidencia de un efecto” no se debería confundir con “evidencia de que no hay efecto”.

Implicaciones para la investigación [fijo, encabezado nivel 2]
Esta sección debería abordar los aspectos clínicos clave que permanecen sin resolver después de la revisión de la evidencia presentada en las revisiones incluidas/subyacentes. Si existen intervenciones potenciales importantes para la afección bajo consideración que no se hayan abordado en una revisión Cochrane de intervenciones, esta laguna se debería señalar claramente en esta sección. Además de proporcionar una agenda para las investigaciones futuras, esta sección puede ser útil para los que toman decisiones clínicas al indicar claramente las áreas en las que aún existe incertidumbre.

Reconocimientos [fijo, encabezado nivel 1]
Esta sección se debería utilizar para reconocer a cualquier persona u organización que los autores deseen reconocer, incluidas las personas que no se enumeran entre los autores (ver Capítulo 4, Sección 4.5).

Contribuciones de los autores [fijo, encabezado nivel 1]
Las contribuciones de los coautores actuales se deberían describir en esta sección (ver Capítulo 4, Sección 4.5).

Declaraciones de interés [fijo, encabezado nivel 1]
Los autores deberían detallar cualquier afiliación presente o pasada u otra relación con cualquier organización o entidad con interés en la revisión, que pudiera dar lugar a un conflicto de interés real o supuesto (ver Capítulo 4, Sección 4.5). Los autores deberían declarar si han participado en algunos de los estudios incluidos en una de las revisiones incluidas, o en la autoría de una revisión sistemática incluida en el Resumen.

Diferencias entre el protocolo y la revisión [fijo, encabezado nivel 1]
En algunas ocasiones es necesario utilizar métodos diferentes de los descritos en el protocolo original (ver Capítulo 4, Sección 4.5).

Notas publicadas [fijo, encabezado nivel 1]
Ver Capítulo 4 (Sección 4.5).

22.3.5 Revisiones y referencias
Los autores deberían verificar la exactitud de todas las referencias.

22.3.5.1 Referencias a las revisiones
Se debería crear un “Identificador de la referencia” para cada revisión incluida y utilizarlo en todo el Resumen. Este generalmente incluye el apellido del autor principal y el año de la cita de la versión más reciente de la revisión (p.ej. Efron 2006). Cuando dos o más revisiones compartan el mismo autor principal y año, se puede agregar una letra (p.ej. Efron 2997a, Efron 2007b). Las revisiones se organizan bajo dos encabezados fijos, como se muestra a continuación.

Revisiones incluidas
Revisiones que cumplen específicamente los criterios de elegibilidad y se incluyen en el Resumen.

Revisiones excluidas
Revisiones (si hay alguna) que no cumplen específicamente los criterios de elegibilidad y no se incluyen en el Resumen.

22.3.5.2 Otras referencias

Se deberían enumerar otras referencias citadas en el texto, incluidas las citadas en las secciones “Antecedentes” y “Métodos”.

22.3.6 Tablas

Se deberían considerar varios tipos de tablas para los Resúmenes; todas se pueden crear como Tablas adicionales en RevMan.

22.3.6.1 Tabla “Características de las revisiones incluidas”

Cada Resumen debería incluir una o más tablas mediante el formato mostrado en la Figura 22.3.a para permitir que los lectores puedan revisar rápidamente los aspectos esenciales de las revisiones Cochrane incluidas en el Resumen.

Notas sobre la compleción de las columnas
Revisión
El “Identificador de la Referencia” para cada revisión incluida (ver Sección 22.3.5.1).

Última actualización evaluada
Esta columna debería enumerar la fecha en la cual la revisión incluida se evaluó por última vez (ver Capítulo 3, Sección 3.2.2). Esta fecha debería estar aproximadamente dentro de los seis meses a partir de la búsqueda de los estudios, y los resultados de esta búsqueda se deberían haber incorporado a la revisión.

Población
Use esta columna para señalar algunas características específicas de la población incluida en la revisión Cochrane, es decir, aquí se debería señalar cualquier limitación en la edad, el sexo, el estadio de la enfermedad, la comorbilidad, etc.

Intervenciones
Enumerar las intervenciones específicas incluidas en el alcance de la revisión, independientemente de si se identificaron y se incluyeron, o no, en la revisión Cochrane estudios con datos relacionados con dichas intervenciones.

Intervenciones de comparación
Enumerar los tipos de intervenciones de comparación que se utilizaron (es decir, placebo, ningún tratamiento o grupos control de intervenciones alternativas).

Resultados para los cuales se proporcionaron datos
Incluir resultados importantes para los cuales la revisión presentó datos, se hayan incluido o no en los datos resumen presentados en el Resumen.

Limitaciones de la revisión
En esta columna, se debería proporcionar una breve descripción de cualquier limitación importante de los métodos utilizados en la revisión Cochrane (o de otro tipo). No utilizar esta columna para resumir la calidad de los estudios identificados en la revisión; dicha información se puede incluir en la tabla “Resumen de las revisiones” (ver Sección 22.3.6.2).

Figura 22.3.a: Plantilla para una tabla "Características de las revisiones incluidas"

Revisión	Última actualización evaluada	Población	Intervenciones	Intervenciones de comparación	Resultados para los cuales se describieron datos	Limitaciones de la revisión

22.3.6.2 Tabla "Resumen de las revisiones"

Cada Resumen debería contener una o más tablas que utilicen el formato mostrado en la Figura 22.3.b para sintetizar los resultados. Este formato de ha diseñado para reflejar (en lo posible) el formato de las tablas "Resumen de los hallazgos" (ver Capítulo 11, Sección 11.5 para orientación adicional). Si el Resumen aborda más de una población clínica (p.ej. grupos que difieren con respecto al estadio o la gravedad de la enfermedad, las comorbilidades u otros factores que es probable que afecten los resultados bajo estudio), entonces se deberían utilizar tablas separadas para las diferentes poblaciones clínicas. No hay dudas de que la forma exacta puede variar con los temas de revisión pero cada tabla debería incluir resultados beneficiosos y perjudiciales, la frecuencia o gravedad de estos resultados en los grupos control, las estimaciones de los efectos relativos y absolutos de las intervenciones, las indicaciones del riesgo de sesgo (que pueden variar para el resultado y la comparación), y cualquier comentario.

Plantilla para una tabla "Resumen de revisiones"
La Figura 22.3.b proporciona una plantilla para una tabla "Resumen de revisiones". La intención es hacer el formato de esta tabla tan similar como sea posible al utilizado para las tablas "Resumen de los hallazgos". Si cambia el formato recomendado para las tablas "Resumen de los hallazgos", el formato recomendado para esta tabla también cambiará.

Los encabezados de las filas
Las filas se deberían organizar por resultado, y comenzar con el resultado primario de interés. Dentro de cada resultado una serie de filas debería proporcionar los resultados de los diferentes pares de intervenciones o comparaciones para las cuales haya datos disponibles. Generalmente se deberán incluir una o más filas para los resultados adversos, incluso si las revisiones incluidas no describieron dichos resultados.

Notas sobre la compleción de las columnas
1.Resultados
Se deberían enumerar los principales resultados beneficiosos y perjudiciales (los más relevantes para los participantes, determinados preferiblemente antes de la compleción de los resultados del Resumen para evitar la posibilidad de seleccionar los resultados descritos según la significación, y no por la importancia clínica). El número de resultados no debería exceder de siete. Los resultados importantes que no tengan datos disponibles también se deberían enumerar en la tabla.

Si se investigan múltiples intervenciones la tabla se debería organizar inicialmente por resultado, con filas incluidas en cada subsección de resultados que presente datos que comparen los resultados de dos intervenciones con respecto al resultado.

2. Riesgo supuesto (con el comparador)
Para cada fila se deberían proporcionar los riesgos representativos del grupo comparador. Los mismos se pueden obtener de los riesgos de los grupos control como se

describen en las revisiones Cochrane incluidas. Si hay variaciones importantes en los riesgos de los grupos control, se deberían incluir dos o tres tasas representativas para cada fila de la tabla que representen una población de bajo, moderado y alto riesgo. Siempre que sea posible, se debería indicar el tipo de participantes a los cuales se puede aplicar un riesgo determinado del grupo control en esta columna, en la columna de comentarios o en las notas al pie.

3. Riesgo correspondiente (con la intervención)
Esta columna tiene como objetivo mostrar el riesgo absoluto esperado de la intervención con uno, dos o tres riesgos comparadores supuestos citados en la columna anterior. Los números se pueden calcular al aplicar el efecto relativo a cada riesgo supuesto para la misma fila (ver Capítulo 11, Sección 11.5.4).

4. Efecto relativo
Para los resultados dicotómicos generalmente se deberían utilizar el cociente de riesgos o el *odds ratio*. En la medida de lo posible la estadística resumen que se utilice se debería estandarizar para todas las revisiones incluidas si diferentes revisiones utilizaron diferentes estadísticas resumen en sus análisis. Se debería incluir el intervalo de confianza del 95% para proporcionar una medida de incertidumbre. Éste se puede calcular mediante el modelo de efectos fijos o el modelo de efectos aleatorios; sin embargo, se debería utilizar el mismo modelo para todos los resultados relacionados con un resultado determinado.

5. Número de participantes y estudios
En muchos casos el número de estudios y participantes para los cuales los datos están disponibles para un resultado y una comparación de tratamientos específicos será menor que el número total de estudios y participantes descritos en la revisión Cochrane de la cual se extrajeron los datos (debido a que la revisión Cochrane puede incluir estudios que no proporcionaron un resultado específico o una comparación específica). De ser así, el número de estudio y participantes descritos en esta columna debería reflejar solamente el subgrupo que proporcionó datos para la comparación y el resultado de interés.

6. Calidad
Comentar sobre la calidad de la evidencia para cada fila de la tabla (es de señalar que, debido a que las diferentes filas pueden contener datos extraídos de diferentes revisiones Cochrane o de diferentes estudios dentro de una revisión Cochrane individual, la calidad de la evidencia puede variar de fila a fila). Se recomienda el uso del sistema de calificación de la evidencia desarrollado por el grupo GRADE (GRADE Working Group 2004), que se incorporó al programa estadístico disponible para los autores de revisiones Cochrane para la preparación de las tablas “Resumen de los hallazgos”. El sistema y los métodos empleados para calificar la calidad de la evidencia se debería describir en la sección “Métodos” del Resumen.

7. Comentarios
El objetivo de este campo es proporcionar comentarios adicionales que ayuden a interpretar la información o los datos identificados en la fila. Por ejemplo, pueden ser sobre al validez de la medida de resultado o la modificación del efecto. Aquí se deberían mostrar las advertencias importantes acerca de los resultados. No todas las filas necesitarán comentarios, por lo que es mejor dejarlo en blanco si no hay comentarios importantes que hacer.

Medidas de resultado continuas
Las medidas de resultado continuas se pueden mostrar en la tabla del Resumen, pero deberían ser clínicamente significativas. Es necesario que las unidades sean claras y fácilmente interpretables, por ejemplo los días de dolor o la frecuencia de cefalea son fácilmente interpretables. Sin embargo, los médicos no especializados y los pacientes no pueden interpretar muchas de las escalas, por ejemplo, los puntos de un Inventario de

Depresión de Beck o las puntuaciones de la calidad de vida. Para éstas, una presentación más significativa pudiera ser expresar los resultados en términos de riesgo (p.ej. una mejoría del 50%), de ser posible, como se discute en el Capítulo 12 (Sección 12.6).

La etiqueta de los resultados también debería ser simple. Por ejemplo, “capacidad para realizar las funciones cotidianas sería preferible a “estado funcional”. Si son necesarios detalles específicos de las definiciones de los resultados, se pueden agregar como notas al pie.

Heterogeneidad

En general la discusión detallada de la heterogeneidad no debería ser parte de la tabla resumen. Sin embargo, si (i) la heterogeneidad da lugar a cambios importantes para la significación clínica o estadística, o (ii) hubo importantes modificadores del efecto, entonces se debería notificar en la columna Comentarios. En ocasiones una modificación importante del efecto puede necesitar una fila o una tabla separadas para describir, por ejemplo, las diferencias en el efecto de la endarterectomía para diferentes grados de estenosis.

Figura 22.3.b: Plantilla de una tabla "Resumen de revisiones"

Intervenciones para la [Afección] en la [Población]							
Resultado	Intervención e intervención de comparación	Riesgos comparativos ilustrativos (IC del 95%)		Efecto relativo (IC del 95%)	Número de participantes (estudios)	Calidad de la evidencia (GRADE)	Comentarios
		Riesgo supuesto	Riesgo correspondiente				
		Con el comparador	Con la intervención				
Resultado #1							
	Intervención/Comparación #1						
	Intervención/Comparación #2						
	Etc...						
Resultado #2							
	Intervención/Comparación #1						
	Intervención/Comparación #2						
	Etc...						
Resultado #3							
	Intervención/Comparación #1						
	Intervención/Comparación #2						
	Etc...						

22.3.6.3 Otras tablas

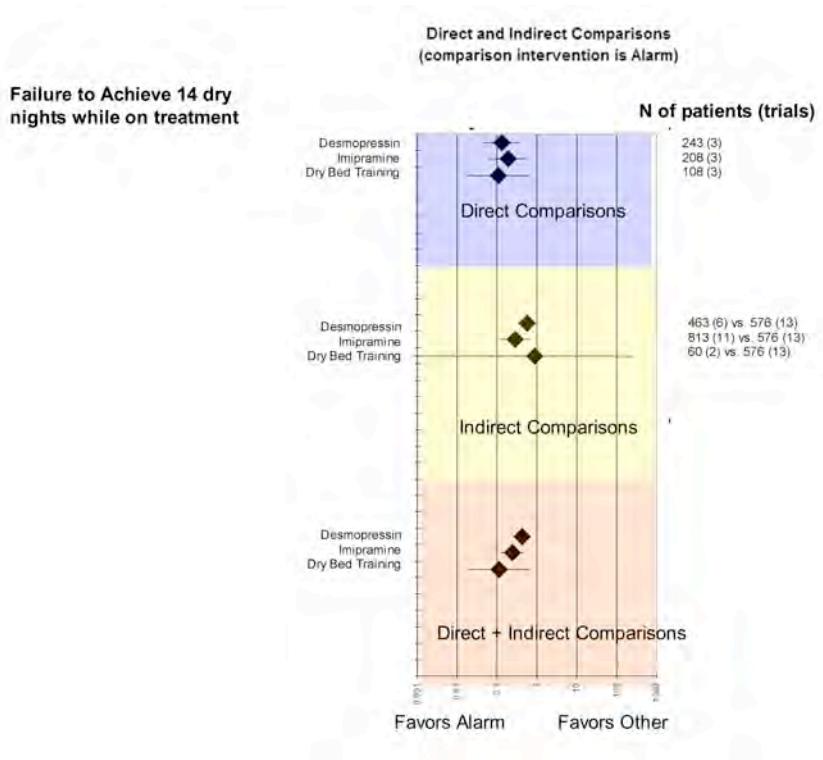
Es posible utilizar otras tablas para la información que no se puede colocar convenientemente en el texto, en las tablas "Características de los estudios incluidos" o en las tablas "Resumen de las revisiones". Se incluyen los siguientes ejemplos:

- Información para apoyar los antecedentes.
- Detalles de los métodos de búsqueda.
- Detalles de las evaluaciones de calidad de las revisiones incluidas.
- Tablas "Resumen de los hallazgos" para las revisiones incluidas, preparadas por los autores del Resumen y que no se encuentran en las revisiones incluidas.

22.3.7 Figuras

La adición de una o dos figuras (como máximo) puede ayudar a los lectores de un Resumen a apreciar mejor las diferencias en la efectividad de las intervenciones que se comparan en la revisión. El formato preferido para las figuras del Resumen es el "diagrama de bosque superior", en el que cada fila de la figura representa los resultados (efecto resumen e intervalo de confianza del 95%) de un metanálisis que compara dos intervenciones. Cada figura debería abordar un resultado único, pero puede incluir varias comparaciones pareadas de intervenciones. Las comparaciones directas, las comparaciones indirectas calculadas y las combinaciones calculadas de comparaciones directas e indirectas se pueden incluir en la misma figura, pero deberían estar claramente identificadas. El texto debería proporcionar información acerca de los métodos utilizados en estos cálculos. En la Figura 22.3.c. se incluye un ejemplo de un diagrama de bosque superior que utiliza datos del Resumen de enuresis (Russell 2006).

Figura 22.3.c: Ejemplo de "diagrama de bosque superior" que compara intervenciones para la enuresis en niños. Este ejemplo se preparó mediante Microsoft Excel



22.4 Información del capítulo

Autores: Lorne A Becker y Andrew D Oxman.

La versión en inglés de este capítulo se debería citar como: Becker LA, Oxman AD. Chapter 22: Overviews of reviews. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Agradecimientos: Los métodos para los Resúmenes Cochrane han sido desarrollados por un grupo de trabajo organizado por el Grupo Directivo de la Colaboración Cochrane, integrado por Lorne Becker (Organizador), Jon Deeks, Paul Glasziou, Jill Hayden, Steff Lewis, Yoon Loke, Lara Maxwell, Andy Oxman, Rebecca Ryan, Denise Thomson, Peter Tugwell y Janet Wale. Se agradece a estas personas por sus contribuciones, y también a Lesley Gillespie, Helen Handoll y Julian Higgins por sus comentarios sobre los borradores anteriores.

22.5 Referencias

Glenny 2005
Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, Bradburn M, Eastwood AJ. Indirect comparisons of competing interventions. *Health Technology Assessment* 2005; 9: 26.

GRADE Working Group 2004
GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004; 328: 1490-1494.

Oxman 1994
Oxman AD. Checklists for review articles. *BMJ* 1994; 309: 648-651.

Petticrew 1999
Petticrew M, Song F, Wilson P, Wright K. Quality-assessed reviews of health care interventions and the database of abstracts of reviews of effectiveness (DARE). NHS CRD Review, Dissemination, and Information Teams. *International Journal of Technology Assessment in Health Care* 1999; 15: 671-678.

Russell 2006
Russell K, Kiddoo D. The Cochrane Library and nocturnal enuresis; an umbrella review. *Evidence-Based Child Health* 2006; 1: 5-8.

Shea 2006
Shea B, Boers M, Grimshaw JM, Hamel C, Bouter LM. Does updating improve the methodological and reporting quality of systematic reviews? *BMC Medical Research Methodology* 2006; 6: 27.