

8. Cómo decidir qué tanta confianza depositar en una revisión sistemática

Simon Lewin^{1}, Andrew D Oxman², John N Lavis³ y Atle Fretheim⁴*

Lewin S, Oxman AD, Lavis JN, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). 8. Deciding how much confidence to place in a systematic review.** Health Research Policy and Systems; 2009, **7(Suppl 1):S8** doi:10.1186/1478-4505-7-S1-S8.

<http://www.health-policy-systems.com/content/pdf/1478-4505-7-S1-s8.pdf>

1. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway;
Health Systems Research Unit, Medical Research Council of South Africa
2. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway
3. Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, and
Department of Political Science, McMaster University, 1200 Main St. West, HSC-2D3, Hamilton, ON,
Canada, L8N 3Z5
4. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway;
Section for International Health, Institute of General Practice and Community Medicine, Faculty of Medicine,
University of Oslo, Norway

* Autor para correspondencia (simon.lewin@nokc.no)

El presente documento es una traducción del artículo publicado en Health Research Policy and Systems, 2009; 7:Supplement 1 (www.health-policy-systems.com/supplements/7/S1). Está permitido su uso, distribución y reproducción ilimitada en cualquier medio, siempre y cuando la fuente esté debidamente citada. En el sitio web de SUPPORT (www.support-collaboration.org) se pueden encontrar enlaces a las traducciones de esta serie al chino, español, francés y portugués. Las sugerencias para correcciones y mejoras de las herramientas de esta serie, que son muy bienvenidas, deben enviarse a: STP@nokc.no.

Esta serie de artículos fue preparada como parte del Proyecto SUPPORT, respaldado por el programa INCO del 6º Programa Marco de la Comisión Europea, contrato 031939. Las siguientes organizaciones financiaron una reunión de revisión por pares donde se discutió una versión anterior de la serie: Norwegian Agency for Development Cooperation (Norad), Alliance for Health Policy and Systems Research (AHPSR) y Milbank Memorial Fund. John Lavis recibió apoyo salarial por su condición de Jefe de Investigación de Canadá en Intercambio y Transferencia de Conocimiento (Canada Research Chair in Knowledge Transfer and Exchange). Las siguientes organizaciones respaldaron la traducción y la disseminación de los artículos: Norad, la sede noruega del Grupo de Revisión Cochrane de Prácticas Efectivas y Organización de la Atención de la Salud (Cochrane Effective Practice and Organisation of Care, EPOC); Norwegian Knowledge Centre for the Health Services, AHPSR, Canadian Health Services Research Foundation (CHSRF), Evidence-Informed Policy Network (EVIPNet) y la Organización Panamericana de la Salud (OPS). Ninguno de los financiadores participaron en la preparación del borrador, la revisión o la aprobación del contenido.

Este artículo fue traducido al español por el Centro Rosarino de Estudios Perinatales (CREP) con el apoyo de Canadian Health Services Research Foundation (CHSRF) www.chsrf.ca, Centro Rosarino de Estudios Perinatales (CREP) www.crep.org.ar, y Organización Panamericana de la Salud (OPS) (www.paho.org/researchportal).



Resumo

El presente artículo forma parte de una serie de artículos destinada a los responsables de la toma de decisiones sobre políticas y programas de salud y a quienes los respaldan.

La confiabilidad de las revisiones sistemáticas en cuanto a los efectos de las intervenciones sanitarias es variable. En consecuencia, los responsables de la toma de decisiones en políticas y las demás personas necesitan evaluar qué tanta confianza puede depositarse en dicha evidencia. El uso de procesos sistemáticos y transparentes para determinar tales decisiones puede ayudar a evitar la aparición de errores y sesgo en estos juicios. En este artículo, sugerimos cinco preguntas que pueden tenerse en cuenta al momento de decidir qué tanta confianza depositar en los hallazgos de una revisión sistemática sobre los efectos de una intervención. Estas son: 1. En la revisión, ¿se abordó explícitamente una política adecuada o una pregunta de conducción? 2. ¿Se utilizaron los criterios adecuados al tener en cuenta los estudios para la revisión? 3. La búsqueda de estudios relevantes, ¿se hizo de manera detallada y razonablemente exhaustiva? 4. ¿Se pudieron reproducir las evaluaciones de la relevancia de los estudios del tema de revisión y de su riesgo de sesgo? 5. ¿Los resultados fueron similares entre los estudios?

ACERCA DE LAS HERRAMIENTAS SUPPORT (STP)

El presente artículo forma parte de una serie de artículos destinada a los responsables de la toma de decisiones sobre políticas y programas de salud y a quienes los respaldan. El objetivo de esta serie de artículos es ayudar a que estas personas se aseguren de tomar las decisiones informándose por la mejor evidencia de la investigación disponible. En la Introducción de esta serie [1], se describen con más detalle las Herramientas SUPPORT y las maneras en las cuales pueden ser utilizadas. Adjunto a cada artículo se encuentra un glosario que se aplica a toda la serie (ver Archivo adicional 1). En el sitio web de SUPPORT (www.support-collaboration.org) se pueden encontrar enlaces a las traducciones de esta serie al español, portugués, francés y chino. Las sugerencias para correcciones y mejoras de las herramientas de esta serie, que son muy bienvenidas, deben enviarse a: STP@nokc.no.

ESCENARIOS

Escenario 1: Usted es un funcionario público de alto rango y va a presentar ante el Ministro una propuesta sobre la evidencia para respaldar una serie de opciones de políticas y programas para abordar un asunto de alta prioridad. Usted está preocupado acerca de qué tanta confianza puede depositarse en las revisiones sistemáticas de la evidencia para cada opción y quiere asegurarse de que éstas hayan sido debidamente evaluadas por su personal.

Escenario 2: Usted trabaja en el Ministerio de Salud y está preparando un documento sobre las opciones para abordar un tema de salud prioritario. Se ha identificado una serie de revisiones sistemáticas de los efectos de las opciones y a usted se le ha solicitado que realice una evaluación acerca de qué tanta confianza puede depositarse en cada revisión.

Escenario 3: Usted trabaja en una unidad independiente que respalda al Ministerio de Salud en el uso de la evidencia para la toma de decisiones en políticas. Usted está preparando un documento para el Ministerio sobre los posibles impactos de las opciones para abordar un tema de salud prioritario. Usted desea asesoramiento para evaluar qué tanta confianza puede depositarse en las revisiones sistemáticas de los impactos de cada opción.

ANTECEDENTES

Para los responsables de la toma de decisiones (Escenario 1), este artículo sugiere una serie de preguntas que podría tener en cuenta su personal al momento de decidir qué tanta confianza depositar en los hallazgos de una revisión sistemática de los efectos de las intervenciones de la atención de la salud.

Para quienes respaldan a los responsables de la toma de decisiones en políticas (Escenarios 2 y 3), este artículo sugiere una serie de preguntas que pueden utilizarse para asesorar una evaluación crítica de revisiones sistemáticas de efectos.

Las revisiones sistemáticas de estudios clínicos controlados aleatorizados cuentan con una amplia aceptación ya que brindan la evidencia más confiable acerca de los efectos de las intervenciones de la atención de la salud [2,3]. Las revisiones sistemáticas se caracterizan por su enfoque sistemático y explícito para identificar, seleccionar y evaluar la investigación relevante, y para recoger y analizar los datos de los estudios incluidos [3]. Las revisiones sistemáticas se están utilizando cada vez más para identificar, evaluar y combinar la evidencia de las consecuencias económicas de las intervenciones [4], tales como la relación costo-efectividad de la promoción de la lactancia para los neonatos de las unidades neonatales [5] o los costos de las estrategias de diseminación e implementación de las distintas guías [6]. También se utilizan para resumir la evidencia de los estudios cualitativos, tales como las opiniones de los consumidores o los prestadores acerca de las intervenciones de salud [7-10]. En este artículo, nos centramos en las revisiones sistemáticas de los efectos de las políticas o programas de atención de la salud. Éstas incluyen revisiones de arreglos de entrega, tales como los efectos de sustituir médicos con enfermeras en la atención primaria [11], y de estrategias para provocar cambios, tales como los efectos de reuniones de educación continua para los profesionales de la salud [12].

El enfoque sistemático y explícito utilizado en una revisión sistemática tiene como objetivo reducir el riesgo de sesgo y los errores que ocurren por el azar, y ayudar a facilitar la evaluación crítica de estas síntesis [13,14]. Sin embargo, el rigor con el que se llevan a cabo estas revisiones sistemáticas es variable. Por lo tanto, las revisiones no son todas igualmente confiables; es decir, las revisiones pueden diferir en cuanto al nivel de confianza que depositamos en sus hallazgos. En consecuencia, confiar simplemente en el hecho de que una evaluación se denomine “revisión sistemática” (o metanálisis) no es suficiente cuando se utilizan hallazgos para informar decisiones en políticas.

Cuando se utilizan revisiones sistemáticas de efectos para informar decisiones en políticas, los responsables de la toma de decisiones en políticas y las demás personas necesitan juzgar qué tanta confianza pueden depositar en esta evidencia. Utilizar un proceso sistemático y transparente puede ayudar a evitar la aparición de errores y sesgos en sus juicios. Un proceso sistemático y transparente además les permite a las otras partes interesadas (*stakeholders*), incluido el público, entender y evaluar estos juicios. Esto es de especial importancia en contextos donde las evaluaciones influyen sobre las recomendaciones o las decisiones sobre intervenciones clínicas o servicios [15], o las decisiones para implementar o interrumpir programas o políticas. La Figura 1 da una idea general de los pasos que implican la búsqueda y evaluación de revisiones sistemáticas para informar a los responsables de la toma de decisiones en políticas.

La confianza en los hallazgos de una revisión sistemática puede verse limitada por varias razones, que incluyen la imposibilidad de:

- Especificar la pregunta y los métodos de la revisión antes de llevar a cabo la revisión en sí, por ejemplo, en un protocolo de revisión publicado
- Especificar un criterio claro de inclusión y exclusión de estudios
- Describir adecuadamente los estudios incluidos en la revisión
- Evaluar el riesgo de sesgo para los estudios incluidos en la revisión
- Evaluar el riesgo de sesgo de publicación, es decir, la posibilidad de que algunos estudios, por lo general aquellos con resultados positivos (“estadísticamente significativos”), tengan más chance que otros de ser publicados y, por lo tanto, de ser incluidos en una revisión.
- Utilizar métodos apropiados para combinar los resultados de los estudios incluidos (en un metanálisis) cuando sea relevante
- Examinar adecuadamente las diferencias en los hallazgos de los estudios incluidos en una revisión (es decir, la “heterogeneidad” de los hallazgos)
- Basar las conclusiones de la revisión en los datos incluidos

Otras posibles limitaciones de las revisiones sistemáticas incluyen conflictos de intereses (que pueden afectar la confiabilidad de una revisión en cualquiera de las formas descritas anteriormente), y las revisiones que se encuentren desactualizadas.

Las variaciones de confiabilidad, por ejemplo, se observaron en un estudio que comparaba la metodología y los componentes de información de las revisiones Cochrane con las revisiones publicadas en revistas impresas. Este estudio descubrió que las revisiones Cochrane incluían componentes que las hacían menos propensas al sesgo. Se descubrió que esta reducción general del riesgo de sesgo en las revisiones Cochrane se debía tanto a las claras descripciones del criterio de inclusión y exclusión, como a la evaluación formal del riesgo de sesgo de los estudios incluidos en cada revisión [16]. De manera similar, otro estudio comparó la calidad metodológica y las conclusiones de las revisiones Cochrane de estudios clínicos de medicamentos con aquellas de revisiones respaldadas por la industria de los mismos medicamentos. Este estudio reveló que las revisiones Cochrane eran más confiables en cuanto a la evaluación de la calidad. Esto se debía a que las revisiones Cochrane consideraban potenciales sesgos con más frecuencia en comparación con revisiones que estaban respaldadas por la industria. Las revisiones respaldadas por la industria también demostraron ser más propensas a recomendar los medicamentos en cuestión sin reservas [17]. Varios estudios de revisiones también informaron diferencias en cuanto a calidad y conclusiones [18-21].

Varias herramientas fueron diseñadas para evaluar la calidad de las revisiones sistemáticas, entre las cuales se encuentran AMSTAR (*A Measurement Tool to Assess Reviews* -Herramienta de Medición para Evaluar Revisiones-) [22], CASP (*Critical Appraisal Skills Programme* -Programa de Habilidades de Evaluación Crítica-) [23], y una desarrollada por Oxman y Guyatt [24] (ver además [25,26]), pero todas contienen

criterios similares. (La herramienta AMSTAR se describe en la Tabla 1.) Varias herramientas incluyen, además, escalas de calificación para darle un puntaje al nivel de confianza que puede depositarse en una revisión. Un creciente número de revisiones ahora incluye dichas evaluaciones. En general, las calificaciones altas indican que es mayor el nivel de confianza que puede depositarse en los hallazgos de las revisiones. Por el contrario, las calificaciones bajas indican que es menor el nivel de confianza que puede depositarse en los hallazgos de las revisiones y que las revisiones deberían ser examinadas en detalle para identificar sus limitaciones más importantes. Sin embargo, deberían observarse tres puntos: En primer lugar, una puntuación o calificación general no indica necesariamente cuáles son los aspectos específicos de una revisión que se llevaron a cabo de manera confiable; algunos pueden haber sido llevados a cabo de manera más confiable que otros. En segundo lugar, el proceso de puntuación en sí mismo involucra ponderar los distintos ítems en la herramienta de evaluación. Puede ser difícil justificar cuáles son los ítems que deben ser más ponderados que otros [27]. Por último, las herramientas de calificación sólo pueden evaluar la confianza de lo que ha sido informado. Cuando no se presenta la información clave acerca de los métodos utilizados en una revisión, puede no estar claro qué se hizo, o la medida en la cual lo que se hizo constituye una limitación importante.

Una *evaluación* acerca de qué tanta confianza debe depositarse en los hallazgos de una revisión necesita diferenciarse de una *interpretación* de los resultados de la revisión en sí misma. La Tabla 2 brinda orientación acerca de qué buscar en los resultados de una revisión de efectos. La orientación para evaluar qué tanta confianza puede depositarse en los hallazgos de las revisiones de estudios cualitativos y de revisiones de estudios económicos se muestra en la Tabla 3.

Una evaluación del grado de confianza que debe depositarse en los hallazgos de las revisiones además necesita diferenciarse de cualquier evaluación que podría hacerse de la *relevancia* de las revisiones sobre preguntas de políticas específicas. Las consideraciones de la relevancia incluyen, por ejemplo, preguntas con respecto a si una revisión brinda evidencia o no de los efectos de las distintas opciones de políticas o programas en estudio, y si los hallazgos de una revisión se aplican al lugar en el cual la política será implementada. El proceso de evaluar la aplicabilidad de los hallazgos de las revisiones sistemáticas se trata en más detalle en el Artículo 9 de esta serie [28].

En este artículo, sugerimos cinco preguntas que pueden tenerse en cuenta al momento de decidir qué tanta confianza depositar en los hallazgos de las revisiones sistemáticas sobre los efectos de las distintas opciones.

PREGUNTAS A TENER EN CUENTA

Las siguientes preguntas pueden orientar a los responsables de la toma de decisiones en políticas al momento de decidir qué tanta confianza depositar en los hallazgos de una revisión sistemática de los efectos de una opción.

1. En la revisión, ¿se abordó explícitamente una política adecuada o una pregunta de conducción?
2. ¿Se utilizaron los criterios adecuados al tener en cuenta los estudios para la revisión?
3. La búsqueda de estudios relevantes, ¿se hizo de manera detallada y razonablemente exhaustiva?
4. ¿Se pudieron reproducir las evaluaciones de la relevancia de los estudios del tema de revisión y de su riesgo de sesgo?
5. ¿Los resultados fueron similares entre los estudios?

1. En la revisión, ¿se abordó explícitamente una política adecuada o una pregunta de conducción?

Un primer paso importante al evaluar qué tanta confianza puede depositarse en los hallazgos de una revisión sistemática es examinar la *pregunta* que se está abordando. El diseño técnico y la conducción de una revisión pueden ser excelentes, pero los hallazgos de una revisión no serán útiles en la toma de decisiones si no han abordado explícitamente una pregunta de política o conducción que sea razonable, adecuada y relevante al tema que está considerando el responsable de la toma de decisiones en políticas.

Una pregunta adecuada de política o conducción deberá:

- *Ser explícita*: En otras palabras, se planteará en detalle en lugar de estar implícita en el material presentado. Si la pregunta de la revisión no fue expresada explícitamente o formulada claramente, es difícil que evalúe de manera adecuada la conducción de la revisión. Esto se debe a que la conducción de la revisión deberá ser considerada, al menos en parte, en relación a la pregunta en sí misma [29]. Por ejemplo, una evaluación acerca de si los criterios utilizados para seleccionar estudios para una revisión fueron adecuados o no, necesita hacerse en relación a la pregunta de la revisión que los estudios tenían como objetivo responder. Además, una pregunta clara ayuda a los lectores a evaluar si una revisión es relevante o no para su trabajo [29]
- *Ser establecida previamente*: En otras palabras, *antes* de que se lleve a cabo la revisión. Es importante que la pregunta de la revisión se especifique antes de que se lleve a cabo la revisión, preferentemente en un protocolo o plan de revisión. Todas las revisiones Cochrane, por ejemplo, están precedidas por un protocolo de revisión publicado y pueden encontrarse ejemplos de estos en la Biblioteca Cochrane, www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME. Si la

pregunta de la revisión no está especificada antes de que se lleve a cabo la revisión, existe un riesgo de que la pregunta pueda ser alterada para adaptarse a la evidencia encontrada, y de esta manera subestimar la confianza en los hallazgos

- *Abordar una pregunta de relevancia para la toma de decisiones en políticas o conducción.* Esto deberá ser evaluado en un contexto específico, basado en la variedad de temas que son importantes en una jurisdicción particular en un momento particular. La pregunta de una revisión puede no ser relevante si:
 - Es demasiado acotada: Por ejemplo, una revisión puede considerar los efectos de un programa sólo sobre un grupo de participantes de edad específica, ubicados en un lugar particular o para un rango restringido de resultados. No sería posible, en este caso, generalizar los resultados a otras poblaciones, lugares u otros resultados
 - Es demasiado amplia: Una revisión, por ejemplo, puede definir un programa que incluye un rango muy amplio de prácticas y no todas estas pueden ser relevantes a una jurisdicción en particular. O una revisión puede plantear una pregunta muy amplia que no es útil desde una perspectiva de toma de decisiones. Una pregunta acerca de si las enfermeras pueden llevar a cabo o no de manera eficiente programas de promoción de la salud, por ejemplo, no será útil al momento de decidir si un cuadro específico de enfermeras, tales como las enfermeras matriculadas, pueden llevar a cabo de manera eficiente un programa de promoción sanitario para un tema de salud específico, como es la prevención del VIH/SIDA
 - No especifica un grupo de comparación adecuado: Si, por ejemplo, un programa se compara con un escenario de “ningún programa” en lugar de compararse con el mejor tratamiento actual para una condición

Una pregunta de revisión bien formulada debería especificar *todo* lo siguiente: Los tipos de población y lugares que cubrirá la revisión (por ej. niños de entre un mes y seis años de edad que vivan en una zona endémica de paludismo); los tipos de programas y comparaciones que se considerarán (por ej. medicamentos antipalúdicos administrados a intervalos regulares -la intervención- comparados con placebo o con ningún medicamento -la comparación-); y los tipos de resultados que son de interés (por ej. paludismo clínico y anemia severa) [30,31]. El acrónimo PICO, que se forma a partir de Población (Population), Intervención (Intervention), Comparación (Comparison), Resultados (Outcomes), se utiliza a veces para resumir los cuatro componentes clave de una pregunta de revisión.

Si bien puede parecer obvia la necesidad de una pregunta de revisión bien formulada, son muchas las revisiones narrativas que fracasan en este aspecto. Una revisión de una muestra de tales revisiones publicadas en importantes revistas médicas especializadas demostró que el 20 % fracasó al plantear su objetivo de manera clara [32].

2. ¿Se utilizaron los criterios adecuados al tener en cuenta los estudios para la revisión?

Los criterios de inclusión y exclusión para una revisión son los listados detallados de los tipos de población, intervenciones, comparaciones y resultados que una revisión tendrá en cuenta. Estos criterios, que se especifican en el protocolo de revisión, determinarán qué estudios estarán incluidos. Por lo tanto, tendrán una fuerte influencia sobre los hallazgos de una revisión, y es importante que estos criterios sean adecuados en relación a la pregunta de la revisión.

Las siguientes preguntas deberían examinarse al momento de evaluar si los criterios utilizados al considerar los estudios para una revisión son apropiados o no.

- ¿La revisión especifica criterios claros de inclusión y exclusión? Estos criterios son importantes como forma de protección contra el sesgo relacionado a la inclusión de estudios en una revisión. Una reciente evaluación de la calidad metodológica de revisiones sistemáticas sobre cirugía general, por ejemplo, descubrió que sólo el 70% de estas revisiones informaron los criterios utilizados para decidir qué estudios incluir en una revisión [18]
- ¿Son los criterios de inclusión y exclusión explícitos en relación con lo siguiente?: Los tipos de población que se incluyen, los tipos de intervenciones y comparaciones que se consideran y los tipos de resultados que se evalúan
- ¿Son los criterios de inclusión y exclusión congruentes con la pregunta de la revisión? [33] Por ejemplo, si el objetivo de una revisión es evaluar la profilaxis y el tratamiento intermitente con antipalúdicos para evitar el paludismo en los niños menores que viven en zonas endémicas, los criterios ¿indican la inclusión de estudios de niños de entornos adecuados y especifican las formas de profilaxis y tratamiento que serán evaluadas? [31] De manera similar, si el objetivo de una revisión es examinar los efectos de las intervenciones para aumentar la proporción de profesionales de la salud que trabajan en zonas rurales y otras zonas marginadas, estos criterios ¿indican el rango de profesionales de la salud que serán incluidos y los tipos de intervenciones educativas y financieras que serán evaluadas? [34]

3. La búsqueda de estudios relevantes, ¿se hizo de manera detallada y razonablemente exhaustiva?

Un aspecto clave de una revisión sistemática es la búsqueda exhaustiva y reproducible de la literatura para los estudios que cumplen con los criterios de elegibilidad de dicha revisión. Este enfoque es uno de los elementos que diferencia a las revisiones *sistemáticas* de las revisiones *narrativas*. La búsqueda sistemática contribuye a minimizar el sesgo en una revisión al asegurar que toda la evidencia relevante sea evaluada. Por lo tanto, ayuda a lograr estimaciones confiables de los efectos de la política o programa que se está examinando [35].

El sesgo de publicación, es decir, la publicación selectiva de los estudios basada en la dirección y fuerza de sus resultados [36], es uno de los modos a través del cual puede

introducirse sesgo en las revisiones. Una reciente revisión examinó hasta dónde la publicación de estudios clínicos aleatorizados se ve influenciada si los resultados hallados eran positivos o no y por la importancia percibida de los hallazgos del estudio clínico. Se demostró que los estudios clínicos con resultados positivos tenían significativamente más probabilidad de ser publicados que los estudios clínicos que presentaban hallazgos negativos [37]. Además, esta revisión y otra investigación demostraron que los estudios clínicos que informaban hallazgos positivos eran publicados antes que otros [38]. Como consecuencia, las revisiones pueden sobreestimar los efectos positivos de los programas, a menos que se intente identificar tanto los estudios publicados como los no publicados.

Las revisiones sistemáticas varían en el grado en que incluyen búsquedas exhaustivas. Una revisión de los informes de las revisiones publicadas sobre el tratamiento del asma, por ejemplo, descubrió que sólo el 52% de las 33 revisiones examinadas incluía una búsqueda razonablemente exhaustiva para la evidencia de los efectos [20]. Por lo tanto, es importante verificar cómo se llevaron a cabo las búsquedas de los estudios relevantes.

Al momento de evaluar si la búsqueda de los estudios relevantes se hizo de manera detallada y razonablemente exhaustiva, deberían examinarse las siguientes preguntas [22].

- ¿La revisión describe en detalle la estrategia utilizada para buscar los estudios relevantes? Este informe debería incluir: 1. La lista de fuentes consultadas, 2. Las palabras clave para buscar estas fuentes (si corresponde) 3. Los años a través de los cuales se hicieron las búsquedas. La Tabla 4 brinda ejemplos de la variedad de las fuentes consultadas en las revisiones publicadas en la Biblioteca Cochrane
- ¿La estrategia de búsqueda incluía bases de datos electrónicas de estudios publicados? Hay disponible una amplia variedad de bases de datos electrónicas de estudios publicados y muchas pueden consultarse sin ningún costo o a un costo muy bajo. Las bases de datos principales incluyen PubMed/MEDLINE (compilada por la Biblioteca Nacional de Medicina, National Library of Medicine, EE.UU.), el Registro Cochrane Central de Estudios Clínicos Controlados (CENTRAL – compilado por la Colaboración Cochrane), y las bases de datos regionales como LILACS (Literatura Latinoamericana y del Caribe en Ciencias de la Salud, Latin American and Caribbean Health Sciences). Los Artículos 4 [39] y 5 [40] de esta serie brindan información adicional para encontrar literatura de investigación relevante
- ¿Las búsquedas de bases de datos electrónicas fueron complementadas por una búsqueda adicional? Esto podría haber incluido un examen de las listas de referencia de los estudios relevantes, contactarse con los autores y expertos en la materia, y la consulta de registros de estudios especializados relacionados con el área del tema de la revisión. Esta búsqueda adicional es útil como un modo de ayudar a identificar tanto los estudios adicionales publicados y no publicados (que

pueden incluir estudios disponibles en la literatura “gris”, es decir en otras fuentes de literatura que no corresponden a las revistas indexadas con revisión por pares)

- ¿Las búsquedas están actualizadas? ¿La revisión especifica el período abarcado por las búsquedas y son estas búsquedas actuales? Una revisión publicada, si bien puede ser relevante para la pregunta de una política, quizás haya utilizado búsquedas que ya tienen varios años. Por lo tanto, es posible que la revisión no incluya la última evidencia relevante y, de esta manera, puede dar una estimación poco confiable de los efectos de la opción de política o programa

4. ¿Se pudieron reproducir las evaluaciones de la relevancia de los estudios del tema de revisión y de su riesgo de sesgo?

Los autores de las revisiones sistemáticas necesitan elaborar dos importantes juicios acerca de cada estudio primario que podría estar incluido en una revisión. En primer lugar, ¿el estudio cumple con los criterios de inclusión de la revisión? En otras palabras, ¿es relevante al tema de la revisión? En segundo lugar, ¿cuál es el riesgo de sesgo en los resultados del estudio? El riesgo de sesgo se refiere al riesgo de “error sistemático, o desviación de la verdad, en los resultados o las inferencias” [27].

También se refiere al tema de que si los resultados de un estudio pueden suponerse como exactos [27]. Debido a que estos juicios afectarán los hallazgos de una revisión, es importante que se presenten de un modo transparente y fácil de reproducir. Otros necesitan poder entender cómo se elaboraron estos juicios y poder repetir estas evaluaciones.

Como se trató anteriormente, las revisiones necesitan especificar criterios claros de inclusión y exclusión para poder protegerse contra sesgo en el proceso de seleccionar estudios para la inclusión. Estos criterios y juicios necesariamente afectarán los hallazgos de la revisión al influir en los estudios seleccionados para su inclusión. El sesgo o los errores en estos juicios pueden minimizarse de las siguientes maneras: En primer lugar, dos revisores deberían decidir de manera independiente qué estudios incluir en una revisión. Las discusiones adicionales con otros revisores también pueden utilizarse para resolver desacuerdos relacionados con la inclusión de un estudio particular. En segundo lugar, las razones para la inclusión de un estudio (y para la exclusión de un estudio que parece relevante) deberían ser registradas en una revisión publicada. Esto les permitirá a los lectores elaborar sus propios juicios con respecto a las decisiones de elegibilidad. Además, le otorga un “registro de auditoría” transparente a la revisión, garantizando de esa manera que el proceso es fácil de reproducir.

La habilidad de una revisión sistemática para alcanzar conclusiones relacionadas con los efectos de una política o programa también depende de la validez de los datos obtenidos de cada estudio incluido. Combinar los resultados de los estudios, o hacer un resumen de los mismos en una revisión, puede arrojar un resultado falso si es poca la validez de los estudios individuales incluidos en la revisión. Por lo tanto, evaluar el riesgo de sesgo en los resultados de los estudios incluidos constituye un elemento

importante de una revisión sistemática. Tales evaluaciones deberían formar parte de la interpretación y las conclusiones de una revisión [27].

Se han desarrollado una serie de distintos enfoques para evaluar la calidad o el riesgo de sesgo para los estudios clínicos controlados [27,41,42]. Si bien no tratamos estos distintos enfoques aquí, es importante observar que las revisiones deberían ser explícitas en relación a los enfoques utilizados y aplicarlos de manera consistente.

Al momento de evaluar la relevancia de los estudios incluidos con el tema de la revisión y el riesgo potencial de sesgo, deberían evaluarse las siguientes preguntas:

- ¿Se utilizó un enfoque explícito y transparente para evaluar la relevancia de los estudios con el tema de la revisión? Una revisión debería afirmar cómo fue evaluada la relevancia y brindar una lista de los estudios incluidos y excluidos
- ¿Se utilizó un enfoque explícito y transparente para evaluar el riesgo de sesgo en los estudios incluidos? Una revisión debería informar la herramienta utilizada para evaluar el riesgo de sesgo, cómo se llevó a cabo la evaluación, y los resultados de la evaluación
- ¿Se tomaron en cuenta los resultados de la evaluación de riesgo de sesgo al interpretar los resultados de una revisión? Cuando el riesgo de sesgo en los estudios incluidos es alto, por ejemplo, podríamos tener menos confianza en los hallazgos de una revisión

5. ¿Los resultados fueron similares entre los estudios?

Los hallazgos de los estudios incluidos en una revisión pueden ser muy parecidos, o pueden variar, en términos de los efectos del programa sobre un resultado en particular. A esta variabilidad entre los estudios incluidos en una revisión generalmente se la conoce como “heterogeneidad” [27]. La variabilidad entre los estudios incluidos en una revisión depende en parte del alcance de la revisión. Si el alcance es amplio, el margen y, por lo tanto, la variabilidad de los estudios incluidos, también podrían ser amplios. Por el contrario, si el alcance de una revisión es limitado, es probable que los estudios incluidos sean similares entre sí.

Si los participantes, las intervenciones o los resultados de los estudios incluidos en una revisión son muy diferentes, esto podría llevar a una variación o a una heterogeneidad si el efecto de la intervención se ve afectado por estos factores. Debido a que el efecto verdadero de la intervención será diferente entre los distintos estudios, en estos ejemplos el efecto promedio entre los estudios no será de mucha ayuda.

Dependiendo del nivel de variabilidad, las revisiones pueden utilizar distintos enfoques para resumir la información de los estudios incluidos, por ejemplo:

- *Calcular el efecto promedio (o combinado) entre los estudios:* Este enfoque es útil cuando la variabilidad entre los estudios es baja. Por ejemplo, una revisión sistemática de los programas de “alta hospitalaria temprana combinada con

hospitalización domiciliaria” (es decir, programas en los cuales los prestadores de salud brindan un tratamiento activo en el hogar del paciente por un tema de salud que de otra manera requeriría una intensa hospitalización para el cuidado del paciente) reveló que los estudios incluidos eran los suficientemente similares para poder estimar el efecto promedio del programa. La revisión reveló evidencia insuficiente de beneficios económicos o de salud de los programas de “alta temprana de hospital combinada con hospitalización domiciliaria” [43]

- *Calcular el efecto promedio para los subgrupos de estudios incluidos en una revisión:* Esto puede resultar útil cuando la variabilidad general de los estudios incluidos en una revisión es alta (y, por lo tanto, no es muy útil para calcular el efecto promedio), pero donde la variabilidad es baja entre los subgrupos de estudios. Por ejemplo, una revisión de intervenciones de trabajadores sanitarios legos en el área de la atención de la salud primaria y de la comunidad agrupó los estudios de acuerdo con los temas de salud tratados por estos trabajadores sanitarios legos. A algunos de los grupos, así como a los trabajadores sanitarios legos que promovieron la inmunización y la lactancia, les fue posible calcular un efecto promedio entre los estudios relevantes. La revisión encontró evidencia que demostró que los trabajadores sanitarios legos pueden mejorar las inmunizaciones y la aceptación de la lactancia [44]
- *Describir el rango de magnitudes de los efectos:* Cuando los estudios no son lo suficientemente similares como para poder calcular un efecto promedio aceptable, aún puede ser posible describir el rango de efectos encontrado en los estudios. Por ejemplo, una revisión de los efectos de auditoría y devolución (*feedback*) sobre la práctica de los prestadores de la atención de la salud demostró que el cumplimiento con la práctica deseada variaba desde una disminución del 16% hasta un aumento del 70%, con una media del 5%. La revisión indicó que la auditoría y devolución (*feedback*) puede mejorar la eficacia de la práctica pero que los efectos son en general pequeños a moderados [45]
- *Catalogar los tipos de intervenciones para tratar un tema en particular:* El amplio alcance de algunas revisiones y, por lo tanto, la variabilidad de los estudios que abarcan, significa que no tiene sentido intentar combinar cuantitativamente los hallazgos de los estudios incluidos, o incluso describir el rango de las magnitudes de los efectos. En estos casos, puede emprenderse una revisión narrativa. Por ejemplo, una revisión sistemática de la efectividad de las intervenciones del servicio de salud para reducir las inequidades en la salud incluyó estudios que evaluaron programas diseñados para reducir las inequidades en la salud y que podrían implementarse dentro del sistema de salud mismo, o en colaboración con otras agencias. El rango de estudios incluidos fue amplio, variando desde programas para mejorar el control de la presión arterial hasta intervenciones de promoción de la salud. Por lo tanto, no se intentó ninguna combinación estadística [46]

En caso de encontrar diferencias en los resultados entre los estudios, deberían evaluarse las siguientes preguntas:

- *¿Existe una explicación convincente para las diferencias que se encontraron? Ésta podría incluir diferencias en los participantes, las intervenciones, los grupos de comparación, los resultados, los lugares o los períodos de tiempo entre los estudios incluidos. Por ejemplo, algunos estudios pueden haber incluido participantes con un rango de edad más amplio o distintas condiciones de salud pre-existentes*
- *Si se realizó una estimación combinada, ¿es probable que ésta sea significativa? Si los estudios incluidos en una revisión son variados, una estimación combinada puede no tener ningún sentido. Puede llevarse a cabo una exploración adicional de los datos, a través de análisis por subgrupos, pero los resultados de dicho análisis exploratorio pueden no ser confiable*

A medida que aumenta el número de revisiones sistemáticas disponibles, se hace más común encontrar más de una revisión sistemática para una pregunta de política en particular. A veces los resultados o las conclusiones de estas revisiones pueden ser diferentes. La Tabla 5 brinda asesoramiento acerca de cómo los responsables de la toma de decisiones en políticas podrían abordar estas situaciones.

CONCLUSIÓN

Las variaciones en el rigor con el que se llevan a cabo las revisiones sistemáticas de efectos son evidentes. Por lo tanto, es importante evaluar la confiabilidad de las revisiones utilizadas para informar las decisiones en políticas, para poder ser capaces de juzgar qué tanta confianza puede depositarse en esta evidencia. Debería utilizarse un enfoque sistemático y transparente para tales evaluaciones y, para esto, se han desarrollado varias herramientas. Sin embargo, estas herramientas sólo pueden utilizarse para evaluar lo que se ha informado. Es por eso que cualquier evaluación que se realice utilizando estas herramientas necesita emprenderse de manera cuidadosa y razonada.

Siempre que la confiabilidad de una revisión sistemática sea baja, los responsables de la toma de decisiones en políticas deberán depositar menos confianza en los hallazgos y ser más prudentes si las utilizan para informar las decisiones en políticas (como se resume en la Figura 2). Cuando los responsables de la toma de decisiones en políticas tomen decisiones basadas en la evidencia presentada en una revisión, deberán considerar las evaluaciones de la confiabilidad de una revisión junto con información adicional, tal como la utilidad de una revisión en relación con la pregunta de política y la evidencia en el contexto local.

RECURSOS

Documentos útiles y lectura complementaria

Higgins JPT, Altman DF: Chapter 8: Assessing risk of bias in included studies. In Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008). Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008. Disponible en: www.cochrane-handbook.org

Counsell C: Formulating Questions and Locating Primary Studies for Inclusion in Systematic Reviews. *Ann Intern Med* 1997, 127: 380-387

Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C et al.: Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007, 7: 10. Disponible en: www.biomedcentral.com/1471-2288/7/10

Enlaces a sitios web

La base de datos Rx for Change:

www.cadth.ca/index.php/en/compus/optimal-ther-resources/interventions – Ésta resume la evidencia de investigación actual acerca de los efectos de las estrategias para mejorar la práctica de prescripción y el uso de medicamentos. Esta base de datos incluye resúmenes, incluidas las evaluaciones de confiabilidad, de revisiones sistemáticas que evalúan los efectos de las estrategias orientadas a los profesionales, la organización de la atención de la salud, y los consumidores.

Grupo de Revisión Cochrane de Prácticas Efectivas y Organización de la Atención de la Salud (Cochrane Effective Practice and Organisation of Care (EPOC) Review Group): www.epoc.cochrane.org/en/index.html – El Grupo Revisor brinda orientación sobre la evaluación de la confiabilidad de los distintos tipos de estudios de efectividad.

Colaboración SUPPORT: www.support-collaboration.org/index.htm – Este proyecto produce resúmenes de revisiones de alta prioridad para países de bajos y medianos ingresos, que incluyen evaluaciones de confiabilidad

CONFLICTOS DE INTERESES

Los autores declaran que no presentan conflictos de intereses.

CONTRIBUCIONES DE LOS AUTORES

SL preparó el primer borrador. ADO, JNL y AF contribuyeron en la elaboración y revisión de este artículo.

AGRADECIMIENTOS

Consulte la Introducción de esta serie para ver los agradecimientos a los financiadores y los colaboradores. Además, nos gustaría agradecerle a Duff Montgomerie por sus valiosos comentarios en una versión anterior de este artículo.

REFERENCIAS

1. Lavis JN, Oxman AD, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). Introduction.** *Health Res Policy Syst.* 2009 7(Suppl 1):I1.
2. Lavis JN, Posada FB, Haines A, Osei E: **Use of research to inform public policymaking.** *Lancet* 2004, 364:1615-21.
3. Oxman AD, Lavis JN, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). 1. What is evidence-informed policymaking.** *Health Res Policy Syst.* 2009 7(Suppl 1):S1
4. Oxman AD, Fretheim A, Lavis JN, Lewin S: **SUPPORT Tools for evidence-informed health Policymaking (STP). 12. Finding and using research evidence about resource use and costs.** *Health Res Policy Syst.* 2009 7(Suppl 1):S12
5. Renfrew MJ, Craig D, Dyson L, McCormick F, Rice S, King SE, Misso K, Stenhouse E, Williams AF: **Breastfeeding promotion for infants in neonatal units: a systematic review and economic analysis.** *Health Technol Assess* 2009, 13:1-iv.
6. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, Whitty P, Eccles MP, Matowe L, Shirran L, *et al*: **Effectiveness and efficiency of guideline dissemination and implementation strategies.** *Health Technol Assess* 2004, 8:iii-72.
7. Carlsen B, Glenton C, Pope C: **Thou shalt versus thou shalt not: a meta-synthesis of GPs' attitudes to clinical practice guidelines.** *Br J Gen Pract* 2007, 57:971-8.
8. Mays N, Pope C, Popay J: **Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field.** *J Health Serv Res Policy* 2005, 10 Suppl 1:6-20.
9. Munro SA, Lewin SA, Smith HJ, Engel ME, Fretheim A, Volmink J: **Patient adherence to tuberculosis treatment: a systematic review of qualitative research.** *PLoS Med* 2007, 4:e238.

10. 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.** *Soc Sci Med* 2005, 61:133-55.
11. Laurant M, Reeves D, Hermens R, Braspenning J, Grol R, Sibbald B: **Substitution of doctors by nurses in primary care.** *Cochrane Database Syst Rev* 2005, 2:CD001271.
12. Forsetlund L, Bjorndal A, Rashidian A, Jamtvedt G, O'Brien MA, Wolf F, Davis D, Odgaard-Jensen J, Oxman AD: **Continuing education meetings and workshops: effects on professional practice and health care outcomes.** *Cochrane Database Syst Rev* 2009, 2:CD003030.
13. Mulrow CD: **Rationale for systematic reviews.** *BMJ* 1994, 309:597-9.
14. Oxman AD, Schunemann HJ, Fretheim A: **Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence.** *Health Res Policy Syst* 2006, 4:20.
15. Oxman AD, Glasziou P, Williams JW, Jr.: **What should clinicians do when faced with conflicting recommendations?** *BMJ* 2008, 337:a2530.
16. Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D: **Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals.** *JAMA* 1998, 280:278-80.
17. Jorgensen AW, Hilden J, Gotzsche PC: **Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review.** *BMJ* 2006, 333:782.
18. Dixon E, Hameed M, Sutherland F, Cook DJ, Doig C: **Evaluating meta-analyses in the general surgical literature: a critical appraisal.** *Ann Surg* 2005, 241:450-9.
19. Jadad AR, Cook DJ, Browman GP: **A guide to interpreting discordant systematic reviews.** *CMAJ* 1997, 156:1411-6.
20. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M, Stevens R: **Systematic reviews and meta-analyses on treatment of asthma: critical evaluation.** *BMJ* 2000, 320:537-40.
21. Linde K, Willich SN: **How objective are systematic reviews? Differences between reviews on complementary medicine.** *J R Soc Med* 2003, 96:17-22.
22. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM: **Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews.** *BMC Med Res Methodol* 2007, 7:10.
23. Critical Appraisal Skills Programme: *10 questions to help you make sense of reviews.* United Kingdom, Public Health Resource. 2006.
24. Oxman AD, Guyatt GH: **Validation of an index of the quality of review articles.** *J Clin Epidemiol* 1991, 44:1271-8.

25. Canadian Coordinating Office for Health Technology Assessment: *Proposed Evaluation Tools for COMPUS*. Ottawa, Ottawa: Canadian Coordinating Office for Health Technology Assessment. 2005.
26. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux L: *Systems to rate the strength of scientific evidence [Evidence report/technology assessment no 47]*. Publication No. 02-E016. Rockville, MD, USA, Agency for Healthcare Research and Quality. 2002.
27. Higgins JPT, Altman DF: **Chapter 8: Assessing risk of bias in included studies**. In *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008)*. Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008.
28. Lavis JN, Oxman AD, Souza NM, Lewin S, Gruen RL, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). 9. Assessing the applicability of the findings of a systematic review**. Health Res Policy Syst. 2009 7(Suppl 1):S9.
29. Counsell C: **Formulating Questions and Locating Primary Studies for Inclusion in Systematic Reviews**. *Ann Intern Med* 1997, 127:380-7.
30. Higgins JPT, Green S: *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 [updated September 2008]*. The Cochrane Collaboration; 2008.
31. Meremikwu MM, Donegan S, Esu E: **Chemoprophylaxis and intermittent treatment for preventing malaria in children**. *Cochrane Database of Systematic Reviews* 2008, 2:CD003756.
32. Mulrow CD: **The medical review article: state of the science**. *Ann Intern Med* 1987, 106:485-8.
33. Oxman AD: **Checklists for review articles**. *BMJ* 1994, 309:648-51.
34. Grobler LA, Marais BJ, Mabunda S, Marindi P, Reuter H, Volmink J: **Interventions for increasing the proportion of health professionals practising in underserved communities**. *Cochrane Database of Systematic Reviews* 2009, 1:CD005314.
35. Lefebvre C, Manheimer E, Glanville J, on behalf of the Cochrane Information Retrieval Methods Group: **Searching for studies**. In *Cochrane Handbook for systematic reviews of interventions. Version 5.0.1 [updated September 2008]*. Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008.
36. Dickersin K, Min YI: **Publication bias: the problem that won't go away**. *Ann N Y Acad Sci* 1993, 703:135-46.
37. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K: **Publication bias in clinical trials due to statistical significance or direction of trial results**. *Cochrane Database Syst Rev* 2009, 1:MR000006.
38. Hopewell S, Clarke M, Stewart L, Tierney J: **Time to publication for results of clinical trials**. *Cochrane Database Syst Rev* 2007, 2:MR000011.
39. Lavis JN, Wilson M, Oxman AD, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). 4. Using research evidence to clarify a problem**. Health Res Policy Syst. 2009 7(Suppl 1):S4.

40. Lavis JN, Wilson MG, Oxman AD, Grimshaw J, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). 5. Using research evidence to frame options to address a problem.** *Health Res Policy Syst.* 2009 7(Suppl 1):S5.
41. Katrak P, Bialocerkowski A, Massy-Westropp N, Kumar VS, Grimmer K: **A systematic review of the content of critical appraisal tools.** *BMC Medical Research Methodology* 2004, 4:22.
42. 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.** *Control Clin Trials* 1995, 16:62-73.
43. Shepperd S, Doll H, Broad J, Gladman J, Iliffe S, Langhorne P, Richards S, Martin F, Harris R: **Early discharge hospital at home.** *Cochrane Database Syst Rev* 2009, 1:CD000356.
44. Lewin SA, Dick J, Pond P, Zwarenstein M, Aja G, van Wyk B, Bosch-Capblanch X, Patrick M: **Lay health workers in primary and community health care.** *Cochrane Database Syst Rev* 2005, 1:CD004015.
45. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD: **Audit and feedback: effects on professional practice and health care outcomes.** *Cochrane Database Syst Rev* 2006, 2:CD000259.
46. Arblaster L, Lambert M, Entwistle V, Forster M, Fullerton D, Sheldon T, Watt I: **A systematic review of the effectiveness of health service interventions aimed at reducing inequalities in health.** *J Health Serv Res Policy* 1996, 1:93-103.
47. Moher D, Jadad AR, Klassen TP: **Guides for reading and interpreting systematic reviews: III. How did the authors synthesize the data and make their conclusions?** *Arch Pediatr Adolesc Med* 1998, 152:915-20.
48. Oxman AD, Cook DJ, Guyatt GH: **Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group.** *JAMA* 1994, 272:1367-71.
49. Oxman AD, Lavis JN, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). 10. Taking equity into consideration when assessing the findings of a systematic review.** *Health Res Policy Syst.* 2009 7(Suppl 1):S10.
50. Oxman AD, Lavis JN, Fretheim A, Lewin S: **SUPPORT Tools for evidence-informed health Policymaking (STP). 17. Dealing with insufficient research evidence.** *Health Res Policy Syst.* 2009 7(Suppl 1):S17.
51. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A: **Synthesising qualitative and quantitative evidence: a review of possible methods.** *J Health Serv Res Policy* 2005, 10:45-53.
52. Noyes J, Popay J, Pearson A, Hannes K, Booth A: **Chapter 20: Qualitative research and Cochrane reviews.** In *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008)*. Edited by Higgins JPT, Green S. The Cochrane Collaboration; 2008.
53. Lavis JN. **Supporting the Use of Systematic Reviews in Policymaking.** *PLoS Med.* In Press.

54. Jefferson T, Demicheli V, Vale L: **Quality of systematic reviews of economic evaluations in health care.** *JAMA* 2002, 287:2809-12.
55. Siegfried N, Muller M, Deeks JJ, Volmink J: **Male circumcision for prevention of heterosexual acquisition of HIV in men.** *Cochrane Database Syst Rev* 2009, 2:CD003362.
56. McGuinness B, Craig D, Bullock R, Passmore P: **Statins for the prevention of dementia.** *Cochrane Database Syst Rev* 2009, 2:CD003160.

Figura 1. Cómo buscar y evaluar revisiones sistemáticas para informar las decisiones acerca de las opciones de políticas y programas

Preguntas		Razones para tener menos confianza en una revisión
¿La revisión aborda explícitamente una pregunta apropiada de política o gestión?	<p>Las revisiones que abordan preguntas poco claras...</p> <p>Una revisión que aborda una pregunta que es diferente de la pregunta de gestión o política que está formulando...</p>	<p>...probablemente ofrezcan respuestas poco claras</p> <p>...puede ser confusa</p>
¿Se utilizaron los criterios adecuados al considerar los estudios para la revisión?	<p>El criterio implícito...</p> <p>Los criterios que no son congruentes con la pregunta que la revisión está abordando...</p>	<p>...puede dar como resultado una selección sesgada de los estudios que deben incluirse</p> <p>...probablemente incluyan los estudios que no son apropiados para responder la pregunta</p>
La búsqueda de estudios relevantes ¿se hizo de manera detallada y razonablemente abarcativa?	<p>Las búsquedas que no están bien diseñadas (y descritas)...</p> <p>Las búsquedas que no son razonablemente abarcativas...</p>	<p>...corren riesgo de no encontrar evidencia relevante de importancia</p> <p>...corren riesgo de estar sesgadas (por ejemplo, tienen mayor probabilidad de encontrar estudios con "resultados positivos") y pueden no encontrar evidencia relevante de importancia</p>
¿Se pudieron reproducir las evaluaciones de la relevancia de los estudios del tema de revisión y de su riesgo de sesgo?	<p>Si no se utilizan criterios explícitos o sólo una persona se ocupa de emitir todos los juicios que son necesarios en una revisión...</p>	<p>...tales juicios pueden no ser reproducibles y pueden estar sesgados</p>
¿Los resultados fueron similares entre los estudios?	<p>Si existen diferencias importantes en los resultados de diferentes estudios y no existe una explicación convincente para tales diferencias...</p>	<p>...es difícil saber cuáles son los verdaderos resultados o qué esperar si se implementa una política similar a las que fueron evaluadas</p>

Figura 2. Los modos en los cuales las revisiones pueden ser poco confiables y confusas

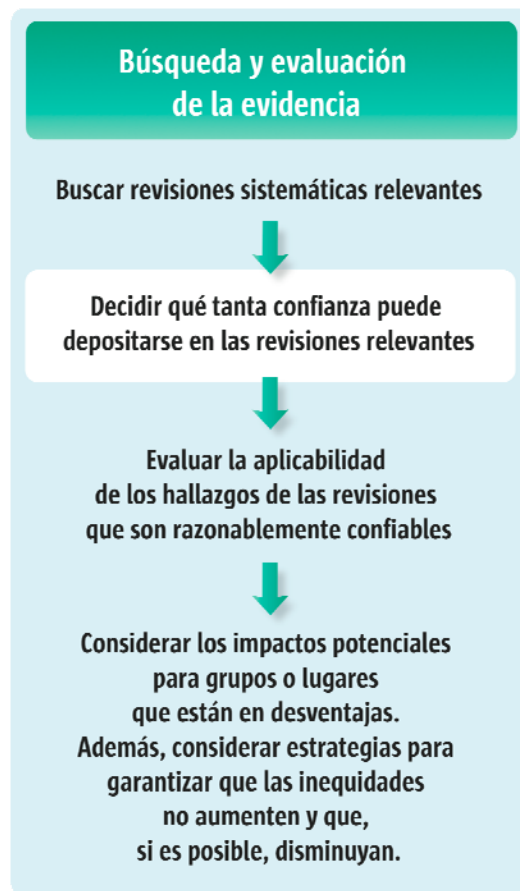


Tabla 1. AMSTAR – Una herramienta de medición para evaluar revisiones ([22])

Para cada pregunta en la herramienta a continuación, los autores de la revisión deben seleccionar una de las opciones “Sí /No /No responde/No corresponde”

1. ¿Se brindó un diseño “a priori”?

La pregunta de la investigación y los criterios de inclusión deberían establecerse antes de llevar a cabo la revisión

2. ¿Hubo duplicación en la selección de estudios y extracción de datos?

Debería haber al menos dos personas independientes a cargo de la extracción de datos, y debería existir un procedimiento consensuado para los desacuerdos

3. ¿Se realizó una búsqueda exhaustiva de literatura?

Deberían consultarse al menos dos fuentes electrónicas. El informe debe incluir los años y las bases de datos utilizadas (por ej. Central, EMBASE y MEDLINE). Deben especificarse las palabras clave y/o los términos MESH y, de ser posible, debe proveerse la estrategia de búsqueda. Todas las búsquedas deberían ser complementadas con consultas a contenidos actuales, revisiones, libros de textos, registros especializados, o expertos en el campo particular de estudio, y mediante la revisión de las referencias en los estudios encontrados

4. ¿Se utilizó el estado de publicación (es decir, literatura gris) como criterio de inclusión?

Los autores deberían especificar que buscaron informes sin tener en cuenta el tipo de publicación. Los autores deberían especificar si excluyeron o no algún informe (de la revisión sistemática), en función del estado de publicación, idioma, etc.

5. ¿Se brindó una lista de estudios (incluidos y excluidos)?

Debería proveerse una lista de estudios incluidos y excluidos

6. ¿Se brindaron las características de los estudios incluidos?

De manera adjunta tal como una tabla, deberían proveerse los datos de los estudios originales sobre los participantes, las intervenciones y los resultados. Deberían informarse los rangos de las características en todos los estudios analizados, por ej, la edad, la raza, el sexo, los datos socioeconómicos relevantes, el estado de enfermedad, la duración, la severidad, o cualquier otra enfermeda

7. ¿Se evaluó y documentó la calidad científica de los estudios incluidos?

Deberían proveerse métodos “a priori” (por ej, para estudios de efectividad si el autor o los autores eligen incluir sólo estudios aleatorizados, de doble ciego, controlados con placebo, u ocultamiento de las asignaciones como criterios de inclusión). Para otros tipos de estudios, serán relevantes los ítems alternativos

8. ¿Se utilizó de manera adecuada la calidad científica de los estudios incluidos al formular las conclusiones?

El rigor metodológico y la calidad científica de los estudios deberían considerarse en el análisis y las conclusiones de la revisión, y plantearse explícitamente al formular las recomendaciones

9. ¿Fueron adecuados los métodos utilizados para combinar los hallazgos de los estudios?

Para los resultados conjuntos, debería hacerse una prueba para garantizar que los estudios pudieran combinarse y para evaluar sus homogeneidad (es decir, la prueba chi-cuadrado para la homogeneidad, I^2). Si existe heterogeneidad debería utilizarse un modelo de efectos aleatorios y/o debería considerarse lo adecuado de la combinación (es decir, ¿fue adecuado combinar los resultados?)

10. ¿Se valoró la probabilidad de sesgo de publicación?

Una evaluación de sesgo de publicación debería incluir una combinación de ayudas gráficas (por ej. un gráfico en embudo – funnel plot – , otras pruebas disponibles) y/o pruebas estadísticas (por ej. prueba de regresión de Egger)

11. ¿Se planteó el conflicto de intereses?

Deberían reconocerse claramente las fuentes posibles de apoyo tanto en la revisión sistemática como en los estudios incluidos

Tabla 2. Cómo interpretar los resultados de las revisiones sistemáticas de efectos

Las siguientes preguntas pueden orientar a los responsables de la toma de decisiones en políticas para interpretar los hallazgos de las revisiones sistemáticas de efectos (adaptado de [33,47,48]*):

- *¿Cuál es la estimación del efecto que se presenta?* Muchas revisiones presentan una estimación promedio del efecto entre los estudios incluidos. Con frecuencia se da en la forma de proporción de riesgo, odds ratio, o diferencia de medias estandarizada
- *¿Es apropiado la estimación promedio del efecto entre los estudios?* Las revisiones utilizan métodos estadísticos para resumir y combinar los datos de los resultados de los estudios incluidos en la revisión. Para asegurarse de que la combinación de los datos de los resultados es apropiada, es útil considerar si los estudios incluidos fueron lo suficientemente similares en términos de población, intervención, comparación y resultados medidos. Cuando no es posible lograr una estimación promedio del efecto, las revisiones en general presentan un resumen narrativo de los datos disponibles
- *¿Se presentaron límites de confianza para la estimación del efecto?* La revisión debería presentar intervalos de confianza alrededor de la estimación promedio del efecto. Cuanto más amplio es el intervalo de confianza, menor es la certeza que podemos tener acerca de la verdadera magnitud del efecto
- *Si se informaron los resultados de los análisis por subgrupos, ¿son estos apropiados?* Una revisión puede presentar hallazgos para un subgrupo particular de participantes entre todos los estudios clínicos o para un subgrupo de estudios [49]. Por ejemplo, una revisión de intervenciones para reducir las enfermedades diarreicas en niños menores de 5 años de edad también podría considerar las intervenciones en niños menores al año de edad. De manera similar, una revisión podría incluir un análisis por subgrupos de estudios que catalogados con un bajo riesgo de sesgo. Un análisis por subgrupos debería tener sentido en relación tanto a la pregunta general de la revisión como al conocimiento previo de factores que podrían haber influido o moderado los efectos de la intervención. Por ejemplo, podría anticiparse que una intervención con una intensidad más alta, podría producir efectos mayores. Los análisis por subgrupos deberían planearse antes de que una revisión se lleve a cabo y debería depositarse menos confianza en estos resultados particulares. Esto se debe a que son menos confiables que los análisis basados en *todos* los estudios clínicos incluidos y porque los análisis estadísticos múltiples pueden producir hallazgos positivos solamente por azar
- *Si hay “no evidenci de efecto”, ¿se toman precauciones para no interpretar esto como “evidencia de no efecto”?* “No evidencia de efecto” no es lo mismo que “evidencia de no efecto”. La primera sugiere que hay disponible evidencia insuficiente como para extraer conclusiones con respecto a los efectos de la

intervención en cuestión. La segunda sugiere que hay clara evidencia en los estudios incluidos de que la intervención *no* tiene los efectos anticipados [50]

- *Las conclusiones y recomendaciones (si hubiere), ¿surgen ambas de la pregunta original de la revisión y la evidencia presentada en la revisión?* Es importante tener en cuenta si las conclusiones presentadas por los autores de la revisión surgen directamente de los datos reunidos de la revisión y no van más allá de esta evidencia
- *¿La evidencia es aplicable a la pregunta de política que se está evaluando?* Las diferencias en los sistemas de salud pueden significar que un programa o intervención que funciona en un entorno puede no funcionar de la misma manera en otro. Los responsables de la toma de decisiones en políticas necesitan evaluar si la evidencia de investigación de una revisión se aplica en su lugar. En el Artículo 9 de esta serie se brinda orientación al respecto [28]

* Hay una superposición entre las preguntas que se enumeran aquí y aquellas destinadas a orientar la evaluación de la confiabilidad de las revisiones sistemáticas. Esto se debe a que la confianza es un elemento importante en la evaluación y comprensión de los resultados de una revisión sistemática

Tabla 3. Cómo evaluar qué tanta confianza puede depositarse en los hallazgos de las revisiones sistemáticas de estudios cualitativos y las revisiones sistemáticas de estudios económicos

Se están llevando a cabo un número cada vez mayor de revisiones sistemáticas de estudios cualitativos. Las mismas utilizan un amplio margen de enfoques, que incluyen síntesis narrativas, metaetnografía y revisión realista. Además de brindar importante información por derecho propio, las revisiones de estudios cualitativos también pueden informar y complementar las revisiones sistemáticas de los efectos [51,52]. Sin embargo, es importante para el lector poder evaluar la confiabilidad de estas revisiones. Hasta la fecha se han diseñado pocas herramientas para este propósito específico. Muchas de las preguntas utilizadas para asesorar a los responsables de la toma de decisiones en políticas al evaluar la confiabilidad de las revisiones sistemáticas de efectos, sin embargo, también son útiles para las revisiones de estudios cualitativos. Éstas incluyen:

1. *En la revisión, ¿se abordó una política adecuada o una pregunta de conducción?*
La pregunta de la revisión debería permitir que se la formule utilizando datos cualitativos y debería ser relevante a la toma de decisiones en políticas. Las revisiones de estudios cualitativos pueden brindar aportes acerca de las opiniones y las experiencias de las partes interesadas con respecto a la salud y a la atención de la salud y de esta manera ayudar a aclarar un problema [39]. Las revisiones de los estudios cualitativos también pueden brindar información acerca de cómo o por qué funcionan las opciones (por ejemplo, a través del examen de evaluaciones de proceso llevadas a cabo a lo largo de la implementación de una política o programa) y acerca de las opiniones de las partes interesadas y sus experiencias relevantes [40,53]
2. *¿Se utilizaron los criterios apropiados para seleccionar los estudios?* La descripción de cómo se seleccionaron los estudios debería ser adecuada en relación con la pregunta de la investigación
3. *¿Se brindó una explicación clara y adecuada para el enfoque de búsqueda utilizado?* Algunas revisiones de estudios cualitativos llevan a cabo búsquedas de literatura exhaustivas mientras que otras utilizan enfoques de muestreo. El enfoque elegido debería estar claramente descrito y justificado
4. *¿Resultó apropiado el enfoque utilizado para evaluar la confiabilidad de los estudios incluidos?* La revisión debería describir cómo se tuvo en cuenta la confiabilidad de los estudios incluidos
5. *¿Se utilizó un enfoque apropiado para analizar los hallazgos de los estudios incluidos?* La revisión debería utilizar un enfoque aceptado para la síntesis y debería describir los fundamentos para el enfoque elegido

Preguntas a tener en cuenta al momento de evaluar la confiabilidad de las revisiones que tienen los estudios económicos (de [54]):

1. ¿Es poco probable que se hayan pasado por alto estudios relevantes importantes?
2. ¿Resultaron apropiados los criterios de inclusión utilizados para seleccionar los artículos?
3. ¿Resultó fácil de reproducir la evaluación de los estudios?
4. ¿Resultaron comparables en términos generales el diseño y/o los métodos y /o el tema de los estudios incluidos?
5. ¿Qué tan fáciles de reproducir son los resultados generales?
6. ¿Contribuirán los resultados a la asignación de recursos en la atención de la salud?

Tabla 4. Ejemplos de fuentes consultadas en las revisiones sistemáticas

Revisión	Fuentes consultadas
<p><i>Revisión de sistemas de salud</i> Ejemplo: Revisión sistemática de las intervenciones de trabajadores sanitarios legos en la atención de la salud primaria y comunitaria</p>	<ol style="list-style-type: none"> 1. Bases de datos electrónicas de estudios publicados: <ul style="list-style-type: none"> • MEDLINE • Registro Cochrane Central de Estudios Clínicos Controlados (CENTRAL) y Registros especializados Cochrane (EPOC y Grupos Revisores de Consumidores y Comunicación) • Science Citations • EMBASE • CINAHL (Cumulative Index to Nursing and Allied Health Literature, Índice Acumulativo de Literatura acerca de Enfermería y Ciencias Aliadas de la Salud) • Healthstar • AMED (Allied and Complementary Medicine Database, Base de datos de Medicina Aliada y Complementaria) • Base de datos de la Eficacia de la Educación de la Salud de Leeds 2. Bibliografías de estudios evaluados para inclusión 3. A todos los autores contactados se les solicitaron estudios adicionales
<p><i>Revisión de salud pública</i> Ejemplo: Revisión sistemática de la circuncisión masculina para la prevención de adquisición de VIH en los hombres [55]</p>	<ol style="list-style-type: none"> 1. Bases de datos electrónicas de estudios publicados: <ul style="list-style-type: none"> • MEDLINE • EMBASE • Registro Cochrane Central de Estudios Clínicos Controlados (CENTRAL) 2. Bases de datos electrónicas de resúmenes de congresos: Bases de datos de Congresos AIDSearch 3. Bases de datos electrónicas de estudios clínicos en curso. <ul style="list-style-type: none"> • ClinicalTrials.gov • Current Controlled Trials 4. Investigadores consultados y organizaciones relevantes en el área 5. Verificadas las listas de referencia de todos los estudios identificados por los métodos anteriormente mencionados y examinadas cualquier revisión sistemática, metanálisis, o guías de prevención identificadas durante el proceso de búsqueda

Revisión	Fuentes consultadas
<p><i>Revisión clínica</i> Ejemplo: Revisión sistemática de las estatinas para la prevención de la demencia [56]</p>	<ol style="list-style-type: none"> 1. Bases de datos electrónicas: <ul style="list-style-type: none"> • El Registro Especializado del Grupo Cochrane de Mejora Cognitiva y Demencia • Registro Cochrane Central de Estudios Clínicos Controlados (CENTRAL) • MEDLINE • EMBASE • PsycINFO (una base de datos de literatura psicológica) • CINAHL • SIGLE (Grey Literature in Europe, Literatura Gris en Europa) • LILACS (Latin American and Caribbean Health Science Literature, Literatura Latinoamericana y del Caribe en Ciencias de la Salud) 2. Bases de datos electrónicas de resúmenes de congresos: <ul style="list-style-type: none"> • ISTP (Index to Scientific and Technical Proceedings, Índice de Procedimientos Científicos y Técnicos) • INSIDE (Bases de datos de la Biblioteca Británica de Actas de Congresos y Revistas) 3. Base de datos electrónicas de tesis: <ul style="list-style-type: none"> • Índice de Tesis (anteriormente ASLIB) (tesis del Reino Unido e Irlanda) • Programa de Tesis Digital de Australia • Tesis y Disertaciones Canadienses • DATAD (Database of African Theses and Dissertations, Base de datos de Tesis y Disertaciones Africanas) • Dissertation Abstract Online (resúmenes en línea de disertaciones) (EE.UU.) 4. Bases de datos electrónicas de estudios controlados en curso: fueron consultadas una gran variedad de dichas bases de datos

Tabla 5. ¿Que deberían hacer los responsables de la toma de decisiones en políticas cuando distintas revisiones sistemáticas que abordan la misma pregunta tienen distintos resultados?

Cuando se busca información para informar una decisión en política particular, es común identificar más de una revisión sistemática relevante. A veces, los resultados de estas revisiones pueden ser diferentes, y esto puede ocasionar que los autores de las revisiones saquen distintas conclusiones acerca de los efectos de una intervención. Este escenario es diferente al escenario en el cual los hallazgos de dos o más revisiones concuerdan pero en el cual los investigadores no están de acuerdo en la interpretación de estos hallazgos [19].

Existen muchas razones por las cuales los resultados de distintas revisiones sistemáticas pueden ser diferentes, que incluyen diferencias en: Las preguntas abordadas por las revisiones, los criterios de inclusión y exclusión utilizados, qué datos se extrajeron de los estudios, cómo se evaluó la calidad de los estudios, y las decisiones con respecto (y los métodos destinados) al análisis estadístico de los datos [19].

La siguiente serie de preguntas diseñadas por Jadad y sus colegas puede utilizarse para orientar la identificación y abordaje de las causas de la discordancia [19].

- ¿Las revisiones abordan la misma pregunta? Si no es así, la revisión elegida debería ser aquella que aborde una pregunta lo más parecida a la pregunta de la política para la cual se necesita evidencia. Por otra parte, debería evaluar los resultados más relevantes a la pregunta de la política
- Si las revisiones abordan la misma pregunta, ¿incluyen los mismos estudios clínicos o estudios primarios? Si no incluyen los mismos estudios clínicos, debería elegirse la revisión que incluya los estudios más relevantes a la pregunta de la política que se está evaluando
- ¿Si las revisiones incluyen los mismos estudios, tienen por lo tanto la misma calidad? Si no es así, debería utilizarse la revisión de mayor calidad

Si ambas revisiones son relevantes, por ejemplo, si las dos abordan aspectos diferentes de la misma pregunta, puede ser útil extraer evidencia de ambas.