

# AMSTAR-2: herramienta de evaluación crítica de revisiones sistemáticas de estudios de intervenciones de salud

AMSTAR-2: critical appraisal tool for systematic reviews of healthcare interventions

Shea BJ y col. BMJ. 2017;358:j4008

## Resumen

El número de revisiones sistemáticas (RS) de estudios de intervenciones sanitarias publicadas ha aumentado rápidamente. Estos documentos de síntesis son ampliamente utilizados para la toma de decisiones clínicas y de políticas de salud, pero están sujetos a una serie de sesgos y es importante que los usuarios puedan identificar las RS de mejor calidad. Considerando la relevancia de la actualización de la herramienta Ameasurement Tool to Assess Systematic Reviews (AMSTAR) para la valoración crítica de RS de estudios de intervenciones en salud, el autor resume los aspectos más relevantes de la publicación de Shea y col., traduce el instrumento y su guía de aplicación, y comenta los aspectos salientes del instrumento AMSTAR-2 junto con sus potenciales implicancias en el desarrollo y reporte de RS.

## Abstract

The number of published systematic reviews (SR) of studies of healthcare interventions has increased rapidly. These synthesis documents are used extensively for clinical and policy decisions, but they are subject to a range of biases and it is important that users can distinguish high quality SR. Taking in consideration the relevance of the updated Ameasurement Tool to Assess Systematic Reviews (AMSTAR) tool for the assessment of SR of healthcare interventions, the author synthesizes the most relevant features of the article by Shea et al, translates the tool and it's user's guide, and comments on the AMSTAR-2 highlights and their potential implications in development and report of SR.

## Principales características del instrumento AMSTAR-2

Se han diseñados muchos instrumentos para evaluar los diferentes aspectos de una revisión, pero pocos de ellos permiten una evaluación crítica integral<sup>1-15</sup>. La herramienta Ameasurement Tool to Assess Systematic Reviews (AMSTAR) fue desarrollado para evaluar RS de ensayos aleatorizados<sup>16-18</sup>. AMSTAR-2 permite una evaluación más detallada de las RS que incluyen también estudios no aleatorizados de intervenciones sanitarias, que son cada vez más incorporados en las RS. De hecho, casi la mitad de las revisiones sistemáticas publicadas incluyen ahora estudios de intervención no aleatorizados (EINA)<sup>19,21</sup>.

AMSTAR-2 es un cuestionario que contiene 16 dominios (ver Apéndice, figura 1), con opciones de respuesta simples: "sí", cuando el resultado es positivo; "no", cuando no se cumplió el estándar o hay información insuficiente para responder; y "sí parcial", en casos en que hubo adherencia parcial al estándar. La herramienta actualizada incluye una extensa guía para el usuario (ver Apéndice, Guía del usuario). Aunque no proporciona una calificación global, de las debilidades en los siete dominios considerados críticos, dado que pueden afectar sustancialmente la validez de una revisión y sus conclusiones (ver cuadro 1), surgen cuatro niveles de confianza: alta, moderada, baja y críticamente baja (ver tabla 1).

Cabe aclarar que no siempre los elementos enumerados en el Cuadro 1 serán considerados como críticos; por ejemplo, el riesgo de sesgo puede ser considerado menos importante cuando una revisión se limita a ensayos controlados aleatorizados (ECA) de alta calidad. Si no se realizó la síntesis estadística de los resultados mediante el meta-análisis, tampoco se aplicará el dominio sobre métodos meta-analíticos.

## Aplicación

Los autores recomiendan que si se empleará AMSTAR-2 para valorar la calidad de una o más RS para basar decisiones importantes a nivel clínico y de políticas sanitarias, el equipo de evaluación acuerde en forma previa ciertas pautas a tener en cuenta al aplicar la herramienta, incluyendo el contexto de la práctica o la política sanitaria y las preguntas clínicas pertinentes (estructura PICO: población, intervención, comparación, resultados). Por ejemplo, las RS disponibles pueden haber incluido estudios con diferentes comparadores o diferentes períodos de seguimiento, por lo que debe establecerse su relevancia para las preguntas importantes relacionadas con la política en cuestión. Las posibles fuentes de sesgo también deben ser acordadas. Por ejemplo, en estudios observacionales de efectos de inter-

Cuadro 1. Dominios críticos de la herramienta AMSTAR-2

1.	Protocolo registrado antes de la revisión (ítem 2)
2.	Adecuada búsqueda en la literatura (ítem 4)
3.	Justificación de los estudios excluidos (ítem 7)
4.	Riesgo de sesgo de los estudios individuales incluidos (ítem 9)
5.	Métodos meta-analíticos apropiados (ítem 11)
6.	Consideración del riesgo de sesgo en la interpretación de los resultados de la revisión (ítem 13)
7.	Evaluación de la presencia y el impacto probable del sesgo de publicación (ítem 15)

Tabla 1. Valoración de la confianza general en los resultados de la revisión

CONFIANZA	JUSTIFICACIÓN
Alta	Ninguna debilidad crítica y hasta una no crítica: la RS proporciona un resumen exacto y completo de los resultados de los estudios disponibles.
Media	Ninguna debilidad crítica y más de una debilidad no crítica (aunque si son muchas podría justificarse una baja confianza): la RS tiene debilidades, pero no hay defectos críticos, pudiendo proporcionar un resumen preciso de los resultados de los estudios disponibles.
Baja	Hasta una debilidad crítica, con o sin puntos débiles no críticos: la RS puede no proporcionar un resumen exacto y completo de los estudios disponibles
Críticamente Baja	Más de una debilidad crítica, con o sin debilidades no críticos: la RS no es confiable

RS: revisión sistemática



venciones, la confusión por indicación (o gravedad de la enfermedad) puede ser problemática cuando las intervenciones se reservan para ciertos subgrupos de pacientes<sup>22</sup>. Asimismo, estos estudios deberían reclutar nuevos usuarios de una tecnología o fármaco para evitar el sesgo de prevalencia<sup>23</sup>. Si el inicio de una intervención tiende a retrasarse, la elección del comparador puede introducir un sesgo de tiempo inmortal<sup>24</sup>. Los errores de medición clasifican erróneamente la exposición y los resultados y pueden desbalancear los grupos de comparación. Una descripción selectiva de múltiples análisis y resultados también pueden desvirtuar los efectos de una intervención. Si bien el manual del usuario de AMSTAR 2 proporciona orientación sobre las distintas secciones (ver Apéndice, Guía del usuario), algunos de los juicios de los evaluadores pueden ser complejos (p.ej. si los autores de la revisión han evaluado adecuadamente el riesgo de sesgo en EINA) y puede necesitarse asesoramiento tanto sobre metodología como sobre el contenido. El conocimiento de los contenidos puede ser a veces necesario para determinar si los autores de la revisión han hecho una evaluación adecuada de los elementos PICO (dominio 1 del AMSTAR-2), e identificar posibles factores de confusión.

#### Acuerdo inter-evaluador

Se midió el acuerdo (mediante kappa) entre evaluadores, quienes tuvieron acceso a la guía del usuario. Los valores variaron sustancialmente a través de los dominios y entre los pares de evaluadores, pero la mayoría de los valores se encontraban en un rango aceptable, con 46 de las 50 puntuaciones kappa en rango de acuerdo moderado o mejor, y 39, acuerdo bueno o mejor.

#### Usabilidad

Los tiempos para completar el AMSTAR-2 oscilaron entre 15 y 32 minutos. Estas estimaciones no incluyen el tiempo necesario para leer la RS. Los comentarios de los evaluadores incluyeron: que se necesita más tiempo para evaluar los análisis de estudios no aleatorizados y mixtos por cuestiones metodológicas de dichos estudios; que era común que los autores de una revisión mencionaran la presencia o ausencia de sesgo de publicación, pero sin aportar ninguna evidencia; y que los autores de la revisión suelen revelar sus posibles conflictos de intereses, pero no la forma en que lo gestionaron.

#### Discusión

Debe remarcar que las respuestas al AMSTAR-2 no deben ser utilizadas para obtener una puntuación global,<sup>25,26</sup> y se recomienda

que los usuarios adopten el proceso de calificación basado en los dominios críticos (Cuadro 1), o en alguna variación sobre la base de estos principios.

Los estudios no aleatorizados grandes, a menudo realizados con grandes bases de datos administrativas, son cada vez más utilizados para evaluar el verdadero impacto mundial de una amplia gama de tecnologías y prácticas de salud. Aunque estos estudios a menudo usan métodos sofisticados, la confusión residual o el fracaso para manejar adecuadamente otras fuentes de sesgo puede conducir a estimaciones de efecto inexactas. La inclusión de grandes estudios observacionales en los meta-análisis puede generar estimaciones precisas, pero sesgadas de los efectos de las intervenciones evaluadas<sup>19</sup>.

Los dominios del AMSTAR-2 provienen de los instrumentos Cochrane de riesgo de sesgo para estudios aleatorizados y no aleatorizados<sup>27,28</sup>. Sin embargo, AMSTAR-2 actualmente no especifica qué instrumentos de riesgo de sesgo deberían haber utilizado los autores de la revisión para evaluar los estudios no aleatorizados incluidos en una RS. El instrumento ROBINS-I, que es la herramienta más completa para los EINA, fue lanzado en 2016, por lo que es poco realista esperar que los autores de las revisiones iniciadas antes de su lanzamiento lo hayan utilizado<sup>28</sup>. Los evaluadores de RS mediante el AMSTAR-2 deberían cerciorarse de que el instrumento para la valoración del riesgo de sesgo utilizado tenga suficiente capacidad discriminativa para los dominios especificados. Una revisión realizada por Sanderson y col. identificó 86 herramientas para evaluar la calidad de los estudios observacionales, sin arribar a conclusiones acerca de si alguna de ellas debería ser preferida sobre las otras<sup>29</sup>. Los instrumentos populares de evaluación para estudios individuales, tales como la escala Newcastle Ottawa y la lista de verificación SIGN no sólo se centran en la validez y además se utilizan para generar una puntuación global, algo que no se recomienda<sup>30,31</sup>.

AMSTAR-2 ofrece una amplia evaluación de la calidad, incluyendo defectos que puedan haber surgido por una mala conducción de la RS (con efectos inciertos en sus resultados). En este sentido, difiere de otro instrumento de riesgo de sesgo en las RS, el ROBIS<sup>32</sup>. ROBIS es un instrumento trifásico sofisticado que se centra específicamente en el riesgo de sesgo introducido por la realización de la RS. Cubre la mayoría de los tipos de preguntas de investigación, incluyendo diagnóstico, pronóstico y etiología. Por el contrario, AMSTAR-2 está destinado a ser utilizado para RS de intervenciones sanitarias, pero inevitablemente hay una superposición de dominios entre ambas.

## Comentario

Aunque no se ha realizado aun una extensa validación de la herramienta AMSTAR-2, ésta sigue los pasos clave en la realización de una RS de manera exhaustiva, reteniendo 10 dominios de la herramienta validada originalmente y adicionando seis más. Además de ser una lista de cotejo para quienes desarrollan o leen una RS, también puede ser muy útil como herramienta de aprendizaje, para valorar uno de los diseños más frecuentemente utilizados en la toma de decisiones basadas en la evidencia. El desarrollo de AMSTAR-2 se basó en gran medida en el consenso de un grupo de expertos, pero también en una amplia retroalimentación de los usuarios del instrumento original y sigue en proceso de refinamiento permanente. De hecho, los autores animan a los usuarios de AMSTAR-2 a proporcionar información si adaptan al instrumento para ámbitos particulares e informar de su experiencia en [www.amstar.ca](http://www.amstar.ca).

En otras herramientas de valoración de RS, incluyendo al AMSTAR original, no se discriminaba la importancia de los ítems, lo que podía generar valoraciones globales engañosas. Una gran fortaleza de esta herramienta es la definición de dominios críticos a partir de los cuales valorar la confianza general en los resultados de la revisión. Podría ser cuestionable que con sólo alguno de los dominios críticos una RS sea considerada de baja confianza. Por ejemplo, la disponibilidad de protocolo, aunque en ascen-

so, no es tan frecuente, y muchas RS bien conducidas podrían ser clasificadas como de baja confianza. A favor de esta posición es que si no existe protocolo, estamos expuestos a innumerable cantidad de decisiones como por ejemplo reportar selectivamente resultados, o realizar análisis post hoc, que podrían sesgar las conclusiones. De hecho, la ausencia de protocolo suele asociarse a RS de menor calidad metodológica. Por otro lado, en el ítem 15 se exige la realización de pruebas gráficas o estadísticas para evaluar la posibilidad de sesgo de publicación. Aunque deseable, esto no siempre es posible cuando el número de estudios es menor a diez. En estos casos no aplicables, si los revisores discutieron apropiadamente la probabilidad y la magnitud del impacto del sesgo de publicación, este ítem debiera considerarse como válido aun sin gráficas o pruebas.

Actualmente desde Cochrane Argentina estamos realizando una prueba de la implementación de esta herramienta en español.

#### Conclusiones del comentarista

AMSTAR-2 ha puesto la vara alta y aunque muchas RS serán clasificadas como de baja confianza, es de esperar que su entrada en escena y futura validación, influyan las nuevas RS y sus actualizaciones, que querrán cumplir con el estándar metodológico presentado.

Agustín Ciapponi. [Centro Cochrane Argentino-Instituto de Efectividad Clínica y Sanitaria (IECS-CONICET) aciapponi@iecs.org.ar]

Ciapponi A. AMSTAR-2: herramienta de evaluación crítica de revisiones sistemáticas de estudios de intervenciones de salud. Evid Act Pract Ambul. 2018;21(1):4-13. Traducido, resumido y comentado de: Shea BJ, y col. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017; 358:j4008. PMID: 28935701 [PubMed - indexed for MEDLINE] PMID: PMC5833365

#### Referencias Bibliográficas

- Whiting P, y col. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol2003;358:25.
- Wong WC, y col. Development of a quality assessment tool for systematic reviews of observational studies (QATSO) of HIV prevalence in men having sex with men and associated risk behaviours. Emerg Themes Epidemiol2008;358:23. PMID:19014686.
- Verhagen AP, y col. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol1998;358:1235-41. PMID:10086815.
- 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. J Epidemiol Community Health1998;358:377-84. PMID:9764259.
- Murray J, y col. Drawing conclusions about causes from systematic reviews of risk factors: The Cambridge Quality Checklists. J Exp Criminol2009;358:1-23.
- Terwee CB, y col. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res2012;358:651-7. PMID:21732199.
- Bérard A, y col. Reliability of Chalmers' scale to assess quality in meta-analyses on pharmacological treatments for osteoporosis. Ann Epidemiol2000;358:498-503. PMID:11118928.
- Thompson S, y col. A proposed method of bias adjustment for meta-analyses of published observational studies. Int J Epidemiol2011;358:765-77. PMID:21186183.
- Deeks JJ, y col. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. Systematic reviews in health care: meta-analysis in context. Wiley & Sons, 2008;285-312.
- Sacks HS, y col. Meta-analyses of randomized controlled trials. N Engl J Med1987;358:450-5. PMID:3807986.
- Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol1991;358:1271-8. PMID:1834807.
- Oxman AD, y col. Evidence-Based Medicine Working Group. Users' guides to the medical literature. VI. How to use an overview. JAMA1994;358:1367-71. PMID:7933399.
- Scottish Intercollegiate Guidelines Network. Methodology Checklist 1: Systematic Reviews and Meta-analyses. www.sign.ac.uk/checklists-and-notes.html
- Aromataris E, y col. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc2015;358:132-40. PMID:26360830.
- Whiting P, y col. ROBIS group. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol2016;358:225-34. PMID:26092286.
- Shea BJ, y col. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol2007;358:10. PMID:17302989.
- Shea BJ, y col. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol2009;358:1013-20. PMID:19230606.
- Shea BJ, y col. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One2007;358:e1350. PMID:18159233.
- Egger M, y col. Spurious precision? Meta-analysis of observational studies. BMJ1998;358:140-4. PMID:9462324.
- Renahan AG, y col. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet2008;358:569-78. PMID:18280327.
- Page MJ, y col. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. PLoS Med2016;358:e1002028. PMID:27218655.
- Salas M, y col. Confounding by indication: an example of variation in the use of epidemiologic terminology. Am J Epidemiol. 1999;358:981-3.
- Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol2003;358:915-20.
- Suissa S. Immortal time bias in pharmaco-epidemiology. Am J Epidemiol2008;358:492-9.
- Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostatistics2001;358:463-71. PMID:12933636.
- Jüni P, y col. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA1999;358:1054-60.
- Higgins JP, y col. Cochrane Bias Methods Group. Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. BMJ2011;358:d5928. PMID:22008217.
- Sterne JA, y col. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ2016;358:i4919.
- Sanderson S, y col. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol2007;358:666-76.
- Wells GA, y col. The Newcastle-Ottawa Scale (NOS) for assessing the quality of observational studies in meta-analyses. www.ohri.ca/programs/clinical\_epidemiology/oxford.asp
- Scottish Intercollegiate Guidelines Network SIGN 50: Methodology Checklist 3: Cohort Studies. www.sign.ac.uk/checklists-and-notes.html
- Whiting P, y col. ROBIS group. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol2016;358:225-34. PMID:26092286
- 225



## POSGRADOS A DISTANCIA

Seguí construyendo tu carrera profesional con nuestros cursos virtuales para profesionales de la salud

- + de 20.000 alumnos en el mundo hispanohablante.
- 12 años de experiencia.
- Equipo docente de reconocida trayectoria en el Hospital Italiano de Buenos Aires.
- Tutores que acompañan y orientan a los alumnos durante toda la cursada.

Encontrá toda la oferta académica en [www.hospitalitaliano.edu.ar](http://www.hospitalitaliano.edu.ar)

Potosí 4234, C.A.B.A. (+5411) 4959-0200 Int. 4519  
Lunes a viernes de 9 a 13 y de 14 a 17 hs.  
campus@hospitalitaliano.org.ar  
www.hospitalitaliano.edu.ar



Instituto Universitario  
Hospital Italiano



**Fig 1. AMSTAR-2:** herramienta de evaluación crítica de revisiones sistemáticas de estudios de intervenciones de salud

1. ¿Las preguntas de investigación y los criterios de inclusión para la revisión incluyen los componentes PICO?

Sí	Opcional	
<input type="checkbox"/> Población	<input type="checkbox"/> Ventana temporal de seguimiento	<input type="checkbox"/> Sí <input type="checkbox"/> No
<input type="checkbox"/> Intervención		
<input type="checkbox"/> Comparación		
<input type="checkbox"/> Resultado (Outcome)		

2. ¿El reporte de la revisión contiene una declaración explícita de que los métodos de la revisión fueron establecidos con anterioridad a su realización y justifica cualquier desviación significativa del protocolo?

Sí Parcial	Sí	
Los autores afirman que tuvieron un protocolo o guía escrita que incluía TODO lo siguiente:	Además de lo anterior, el protocolo debe estar registrado y también debería haber especificado:	<input type="checkbox"/> Sí <input type="checkbox"/> Sí Parcial <input type="checkbox"/> No
<input type="checkbox"/> Pregunta(s) de la revisión	<input type="checkbox"/> Un meta-análisis / plan de síntesis, si aplicara, y	
<input type="checkbox"/> Una estrategia de búsqueda	<input type="checkbox"/> Un plan para investigar causas de heterogeneidad	
<input type="checkbox"/> Criterios de inclusión / exclusión	<input type="checkbox"/> Justificación para cualquier desviación del	
<input type="checkbox"/> Evaluación del riesgo de sesgo		

3. ¿Los autores de la revisión explicaron su decisión sobre los diseños de estudio a incluir en la revisión?

Para sí, la revisión debe satisfacer UNA de las siguientes opciones:	
<input type="checkbox"/> Explicación para incluir sólo Ensayos Clínicos Aleatorizados (ECA), o	<input type="checkbox"/> Sí <input type="checkbox"/> No
<input type="checkbox"/> Explicación para incluir sólo Estudios No Aleatorizados de Intervención (EINA), o	
<input type="checkbox"/> Explicación para incluir ambos: ECA y EINA	

4. ¿Los autores de la revisión usaron una estrategia de búsqueda bibliográfica exhaustiva?

Para sí parcial (TODO lo siguiente):	Para sí, también debería tener (TODO lo siguiente):	
<input type="checkbox"/> Buscaron por lo menos en 2 bases de datos (relevantes a la pregunta de investigación)	<input type="checkbox"/> Haber buscado en listas de referencias / bibliografía de los estudios incluidos	<input type="checkbox"/> Sí <input type="checkbox"/> Sí Parcial <input type="checkbox"/> No
<input type="checkbox"/> Proporcionaron palabras clave y/o estrategia de búsqueda	<input type="checkbox"/> Haber buscado en registros de ensayos/estudios	
<input type="checkbox"/> Explicitan si hubo restricciones de publicación y está justificada (por ejemplo, idioma)	<input type="checkbox"/> Haber incluido o consultado expertos en el campo de estudio	
	<input type="checkbox"/> Haber buscado literatura gris, si correspondiese	
	<input type="checkbox"/> Haber realizado la búsqueda dentro de los 24 meses de finalizada la revisión protocolo	

5. ¿Los autores de la revisión realizaron la selección de estudios por duplicado?

Para sí, UNA de las siguientes:	
<input type="checkbox"/> Al menos dos revisores estuvieron de acuerdo de forma independiente en la selección de los estudios elegibles y consensuaron qué estudios incluir, o	<input type="checkbox"/> Sí <input type="checkbox"/> No
<input type="checkbox"/> Dos revisores seleccionaron una muestra de los estudios elegibles y lograron un buen acuerdo (al menos 80%), siendo el resto seleccionado por un solo revisor	

6. ¿Los autores de la revisión realizaron la extracción de datos por duplicado?

Para sí, UNA de las siguientes:	
<input type="checkbox"/> Al menos dos revisores alcanzaron consenso sobre los datos a extraer, o	<input type="checkbox"/> Sí <input type="checkbox"/> No
<input type="checkbox"/> dos revisores extrajeron los datos de una muestra de los estudios elegibles y lograron un buen acuerdo (al menos 80%), siendo el resto extraído por un solo revisor	

7. ¿Los autores de la revisión proporcionaron una lista de estudios excluidos y justificaron las exclusiones?

Para sí parcial (TODO lo siguiente):	Para sí, también describe (TODO lo siguiente):	
<input type="checkbox"/> Se proporciona una lista de todos los estudios potencialmente relevantes, evaluados por texto completo, pero excluidos de la revisión	<input type="checkbox"/> Fue justificada la exclusión de la revisión de cada estudio potencialmente relevante	<input type="checkbox"/> Sí <input type="checkbox"/> Sí Parcial <input type="checkbox"/> No

8. ¿Los autores de la revisión describieron los estudios incluidos con suficiente detalle?

Para sí parcial (TODO lo siguiente):	Para sí, también describe (TODO lo siguiente):	
<input type="checkbox"/> Poblaciones	<input type="checkbox"/> Población en detalle	<input type="checkbox"/> Sí <input type="checkbox"/> Sí Parcial <input type="checkbox"/> No
<input type="checkbox"/> Intervenciones	<input type="checkbox"/> Ámbito del estudio	
<input type="checkbox"/> Comparadores	<input type="checkbox"/> Marco temporal para el seguimiento	
<input type="checkbox"/> Resultados	<input type="checkbox"/> Intervención y comparador en detalle (incluidas dosis si fuese pertinente)	
<input type="checkbox"/> Diseños de investigación		

9. ¿Los autores de la revisión usaron una técnica satisfactoria para evaluar el riesgo de sesgo de los estudios individuales incluidos en la revisión?

Ensayos Clínicos Aleatorizados (ECA)		
Para sí parcial debe haber valorado:	Para sí, también debe haber valorado:	<input type="checkbox"/> Sí <input type="checkbox"/> Sí Parcial <input type="checkbox"/> No <input type="checkbox"/> Sólo incluye EINA
<input type="checkbox"/> Enmascaramiento de la asignación, y	<input type="checkbox"/> Generación de la secuencia aleatoria, y	
<input type="checkbox"/> cegamiento de pacientes y evaluadores de resultados (innecesario para resultados objetivos como mortalidad por todas las causas)	<input type="checkbox"/> reporte selectivo entre múltiples medidas o análisis de resultados específicos	
Estudios No Aleatorizados de Intervención (EINA)		
Para sí parcial debe haber valorado:	Para sí, también debe haber valorado:	<input type="checkbox"/> Sí <input type="checkbox"/> Sí Parcial <input type="checkbox"/> No <input type="checkbox"/> Sólo incluye EINA
<input type="checkbox"/> Sesgo de confusión, y	<input type="checkbox"/> Métodos utilizados para determinar exposiciones y resultados, y	
<input type="checkbox"/> sesgo de selección	<input type="checkbox"/> reporte selectivo entre múltiples medidas o análisis de resultados específicos	

10. ¿Los autores de la revisión reportaron las fuentes de financiación de los estudios incluidos en la revisión?

Para sí:	
<input type="checkbox"/> Debe haber informado sobre las fuentes de financiación para los estudios individuales incluidos en la revisión Nota: informar que los revisores buscaron esta información pero que no fue reportado por los autores del estudio, también califica	<input type="checkbox"/> Sí <input type="checkbox"/> No

11. Si se realizó un meta-análisis, ¿los autores de la revisión usaron métodos apropiados para la combinación estadística de resultados?

Ensayos Clínicos Aleatorizados (ECA)		
Para sí:		<input type="checkbox"/> Sí <input type="checkbox"/> No <input type="checkbox"/> No Meta-Análisis
<input type="checkbox"/> Los autores justifican la combinación de los datos en un meta-análisis, y		
<input type="checkbox"/> utilizaron una técnica apropiada de ponderación para combinar los resultados de los estudios, ajustada por heterogeneidad si estuviera presente, e		
<input type="checkbox"/> investigaron las causas de la heterogeneidad		
Ensayos Clínicos Aleatorizados (ECA)		
Para sí:		<input type="checkbox"/> Sí <input type="checkbox"/> No <input type="checkbox"/> No Meta-Análisis
<input type="checkbox"/> Los autores justifican la combinación de los datos en un meta-análisis, y		
<input type="checkbox"/> utilizaron una técnica apropiada de ponderación para combinar los resultados de los estudios, ajustada por heterogeneidad si estuviera presente, y		
<input type="checkbox"/> combinaron estadísticamente las estimaciones de efecto de EINA que fueron ajustados por confusión, en lugar de combinar datos crudos, o justificaron combinar datos crudos las estimaciones de efecto ajustado cuando no hubieran estado disponibles, y		
<input type="checkbox"/> reportaron estimaciones de resumen separadas para los ECA y EINA por separado cuando ambos se incluyeron en la revisión		



12. Si se realizó un meta-análisis, ¿los autores de la revisión evaluaron el impacto potencial del riesgo de sesgo en estudios individuales sobre los resultados del meta-análisis u otra síntesis de evidencia?

Para sí:	<input type="checkbox"/> Sí
<input type="checkbox"/> Sólo incluyeron ECA de bajo riesgo de sesgo, o	<input type="checkbox"/> No
<input type="checkbox"/> Si la estimación combinada se basó en ECA y/o EINA con diferentes riesgos de sesgo, los autores realizaron análisis para investigar su posible impacto en las estimaciones sumarias del efecto	<input type="checkbox"/> No Meta-Análisis

13. ¿Los autores de la revisión consideraron el riesgo de sesgo de los estudios individuales al interpretar / discutir los resultados de la revisión?

Para sí:	<input type="checkbox"/> Sí
<input type="checkbox"/> Sólo incluyeron ECA de bajo riesgo de sesgo, o	<input type="checkbox"/> No
<input type="checkbox"/> Si se incluyeron ECA con moderado o alto riesgo de sesgo, o EINA, la revisión proporcionó una discusión sobre el probable impacto de los riesgos de sesgo en los resultados.	

14. ¿Los autores de la revisión proporcionaron una explicación satisfactoria y discutieron cualquier heterogeneidad observada en los resultados de la revisión?

Para sí:	<input type="checkbox"/> Sí
<input type="checkbox"/> No hubo heterogeneidad significativa en los resultados, o	<input type="checkbox"/> No
<input type="checkbox"/> Si hubo heterogeneidad, los autores realizaron una investigación de sus fuentes y discutieron su impacto en los resultados de la revisión.	

15. Si se realizó síntesis cuantitativa ¿los autores de la revisión llevaron a cabo una adecuada investigación del sesgo de publicación (sesgo de estudio pequeño) y discutieron su probable impacto en los resultados de la revisión?

Para sí:	<input type="checkbox"/> Sí
<input type="checkbox"/> Realizaron pruebas gráficas o estadísticas para sesgo de publicación y discutieron la probabilidad y la magnitud del impacto del sesgo de publicación	<input type="checkbox"/> No
	<input type="checkbox"/> No Meta-Análisis

16. ¿Los autores de la revisión informaron de cualquier fuente potencial de conflicto de intereses, incluyendo cualquier financiamiento recibido para llevar a cabo la revisión?

Para sí:	<input type="checkbox"/> Sí
<input type="checkbox"/> Los autores informaron carecer de conflicto de intereses, o	<input type="checkbox"/> No
<input type="checkbox"/> Los autores describen sus fuentes de financiación y cómo fueron gestionados los potenciales conflictos de intereses.	

## Guía del usuario AMSTAR-2

Muchos de los dominios en AMSTAR-2 resultan auto-explicativos. Sin embargo, se brinda una orientación adicional dado que a menudo los temas son complejos y están sujetos a diferentes interpretaciones, en un amplio espectro de intervenciones.

Los criterios deben ser aplicados en entornos donde se realizan revisiones sistemáticas (RS) de intervenciones (generalmente clínicas) bien definidas.

AMSTAR 2 no está diseñado para generar un puntaje global ya que son los dominios críticos los que más pueden debilitar la confianza en una revisión sistemática.

### 1. ¿Las preguntas de investigación y los criterios de inclusión para la revisión incluyen los componentes PICO?

Es una práctica común usar la descripción PICO (población, inter-

vención, grupo de control y resultado) como marco organizativo para una pregunta de un estudio. Algunas veces debe adicionarse el marco temporal si esto es crítico para determinar la probabilidad de que un estudio capte resultados clínicos relevantes (por ejemplo, si un efecto de la intervención solo se espera después de varios años). Los componentes PICO deberían describirse detalladamente en el reporte de la RS y deberían permitir que el evaluador juzgue la selección de estudios, su "combinabilidad" y la aplicabilidad de los resultados. Los autores de las RS no siempre explicitan los elementos de PICO, pero estos deben ser discernibles a través de una lectura cuidadosa de las secciones del resumen, la introducción y los métodos. Para calificar 'Sí' los evaluadores deben asegurarse que los 4 elementos PICO se describen en alguna parte del informe.

## 2. ¿El reporte de la revisión contiene una declaración explícita de que los métodos fueron establecidos con anterioridad a su realización y se justifica cualquier desviación significativa del protocolo?

Las RS son una forma de investigación de tipo observacional, y los métodos para realizarla deberían acordarse antes de que comience la revisión. La adhesión a un protocolo bien desarrollado reduce el riesgo de sesgo en la revisión. Los autores deben demostrar que trabajaron con un protocolo escrito con verificación independiente. Esto puede tomar la forma de registro (por ejemplo, en PROSPERO, <https://www.crd.york.ac.uk/PROSPERO/>), una revista de publicación abierta (por ejemplo, BMJ open) o una fecha de presentación a una oficina de investigación o comité de ética de protocolos de investigación. Las preguntas de investigación y los métodos deberían haberse planificado antes de realizar la revisión. Como mínimo, esto debería indicarse en el reporte (puntaje 'Sí parcial'). Para calificar "Sí", los autores deberían demostrar que trabajaron con un protocolo escrito con verificación independiente antes de que se realizara la revisión, y cuando el protocolo registrado esté disponible, los evaluadores deben compararlo con el informe publicado. Si hay desviaciones, los evaluadores deberían determinar si estas fueron reportadas y justificadas por los autores. Las discrepancias obvias e inexplicadas deberían bajar la calificación.

## 3. ¿Los autores de la revisión fundamentaron su decisión sobre los diseños de estudio a incluir?

La selección de los tipos de estudio a incluir en las RS no debe ser arbitraria. Los autores deben indicar que estrategia siguieron. La regla general (esto puede tener que inferirse de lo que los autores realmente escribieron) es que primero se preguntaran si una revisión restringida a ensayos clínicos aleatorizados (ECA) habría dado un resumen incompleto de los efectos de un tratamiento. Esto podría resultar de la ausencia de ECA relevantes o por la ausencia de resultados en los ECA disponibles (por lo general ocurre con los resultados vinculados a los daños para los pacientes), por un poder estadístico inadecuado, por poblaciones restrictivas o tratamientos de control o intervenciones no representativas. Si la respuesta a esta pregunta general es sí, la inclusión de estudios no aleatorizados de intervención (EINA) está justificada. Por el contrario, para justificar la restricción de la revisión a ECA, los autores deberían argumentar que pueden proporcionar un cuadro completo de los efectos de interés. La restricción de una revisión solo a EINA está justificada cuando los ECA no pueden proporcionar los datos de resultado necesarios, o en el caso en que se hayan completado las revisiones de ECA, y la revisión de EINA complementará lo que ya se conoce. La inclusión de ambos, ECA y EINA, puede justificarse para obtener un panorama completo de la efectividad y los daños asociados con una intervención en esta situación. Recomendamos (ver más abajo) que estos dos tipos de estudios sean evaluados y combinados en forma separada (si el meta-análisis es apropiado). Esta es un área bastante descuidada, e incluso con orientación, puede ser difícil juzgar en qué medida una revisión cumple con los criterios de clasificación. Es posible que deba inferirse la justificación para la selección de diseños de estudio a partir de una lectura cuidadosa del reporte completo del estudio.

## 4. ¿Los autores de la revisión usaron una estrategia de búsqueda bibliográfica exhaustiva?

Deben buscarse al menos en dos bases de datos bibliográficas. El informe debe incluir los años y las bases de datos examinadas (p. ej., Central, EMBASE y MEDLINE). Debe reportarse las palabras clave y / o términos MESH, y la estrategia de búsqueda completa debe estar disponible ante su solicitud. Las búsquedas deben complementarse con el análisis de las revisiones publicadas, los regis-

tros especializados y la consulta con expertos en el tema de estudio, y revisando la lista de referencias de los estudios encontrados. A veces es necesario contactar a los autores de los estudios originales para aclarar los resultados u obtener actualizaciones o correcciones. Se deben buscar publicaciones en todos los idiomas relevantes, y se debe proporcionar una justificación cuando haya restricciones de idioma. Hemos destacado la necesidad de buscar en la literatura gris en algunos casos. A veces la literatura gris es importante, como ocurre con informes de políticas y evaluaciones de programas que solo están disponibles en sitios web (por ejemplo, del gobierno o de agencias de evaluación de tecnologías sanitarias, que pueden o no haber estado sujetos a revisión por pares. Cuando la literatura gris se considere importante, los autores deberían haber buscado fuentes apropiadas, como registros de ECA, resúmenes de conferencias, disertaciones e informes no publicados en sitios web personales (por ejemplo, universidades, redes sociales como ResearchGate). Además, los ensayos de intervenciones médicas que no se hayan publicado en revistas revisadas por pares, pueden obtenerse directamente de las compañías patrocinantes o directamente de los investigadores. Para calificar "Sí" los evaluadores deben estar satisfechos de que todos los aspectos relevantes de la búsqueda hayan sido abordados por los autores de la revisión.

## 5. ¿Los autores de la revisión realizaron la selección de estudios por duplicado?

La mejor práctica requiere dos revisores para determinar la elegibilidad de los estudios para su inclusión en la RS. Esto implica verificar las características de cada estudio (por título, resumen y texto completo) contra los componentes de la pregunta de investigación. Debería haber habido al menos dos asesores independientes para la selección de estudios. En las opciones de respuesta, señalamos la conveniencia de que los autores de la revisión describan el acuerdo inter-evaluador en una muestra de estudios elegibles. Debería haberse utilizado un proceso de consenso cuando surgieron desacuerdos en la selección de los estudios. En el caso de que una sola persona seleccionara los estudios, un segundo revisor debería haber verificado el acuerdo sobre una muestra representativa de estudios y deberían haber logrado un puntaje kappa fuerte (mayor o igual de 0,80).

## 6. ¿Los autores de la revisión realizaron la extracción de datos por duplicado?

Al igual que en el punto 5, debería haber habido al menos dos evaluadores independientes para la extracción de datos, y es necesario un proceso de consenso para cuando surgieron desacuerdos. En el caso que un solo individuo haya extractado los datos, un segundo revisor debería haber verificado el grado de acuerdo en una muestra representativa de estudios y deberían haber alcanzado un puntaje kappa (mayor o igual de 0,80).

## 7. ¿Los autores de la revisión proporcionaron una lista de estudios excluidos y justificaron las exclusiones?

Los autores de la revisión deben proporcionar una lista completa de estudios potencialmente relevantes y la justificación para la exclusión de cada uno. La no inclusión de estudios puede deberse a una variedad de razones, basadas en poblaciones, intervenciones y controles inapropiados o irrelevantes. La exclusión no debe basarse en el riesgo de sesgo, que se trata por separado y posteriormente en el proceso de revisión. La exclusión injustificada puede sesgar los hallazgos de la revisión, y es preferible un enfoque inclusivo en las primeras etapas de una revisión. Este dominio requiere que los autores de la revisión proporcionen una lista completa de estudios potencialmente relevantes con justificación para la exclusión de cada uno.



## 8. ¿Los autores de la revisión describieron los estudios incluidos con suficiente detalle?

Se deben proporcionar la descripción de los temas, las intervenciones, los controles, los resultados, el diseño, el análisis y el ámbito de los estudios incluidos en la revisión. Los detalles deberían ser suficientes para que un evaluador o usuario juzgue en qué medida los estudios fueron elegidos de manera apropiada (en relación con la pregunta PICO) y si las poblaciones e intervenciones en estudio son relevantes para su práctica o la política de salud en cuestión. Las descripciones también proporcionan un marco para estudiar la heterogeneidad en los efectos de la intervención (por ejemplo, por dosis, rango de edad, entorno clínico, etc.).

## 9. ¿Los autores de la revisión usaron una técnica satisfactoria para evaluar el riesgo de sesgo de los estudios individuales incluidos en la revisión?

Esta es una parte crucial de la evaluación de cualquier RS, particularmente aquellas que incluyen EINA. La pregunta clave de evaluación es si en la revisión se tuvo en cuenta el riesgo de sesgo al resumir e interpretar los resultados.

Cuando la revisión se limita a ECA, recomendamos consultar el Manual Cochrane para determinar si los autores de la revisión han hecho una adecuada evaluación de riesgo de sesgo de los ECA individuales. Esta sección se refiere al desafío planteado por los EINA. Los autores de la revisión deberían haber utilizado un enfoque sistemático para la evaluación de riesgo de sesgo, preferiblemente con un instrumento de calificación adecuadamente desarrollado. Si se ha usado un instrumento no estándar, éste debería ser capaz de detectar fallas metodológicas serias. Hay varios instrumentos de riesgo de sesgo (para estudios individuales), incluida la escala de Newcastle Ottawa, la escala SIGN, y la Herramienta de Evaluación de Métodos Mixtos (MMAT). El instrumento de evaluación más completo es el instrumento Cochrane presentado recientemente, ROBINS-I, por lo que es posible que esta herramienta no haya estado disponible en el momento en que se realizó una revisión. Los desarrolladores de AMSTAR 2 nos basamos en los instrumentos Cochrane de riesgo de sesgo para ECA

([http://handbook.cochrane.org/chapter\\_8/8\\_5\\_the\\_cochrane\\_collaborations\\_tool\\_for\\_assessing\\_risk\\_of\\_bias.htm](http://handbook.cochrane.org/chapter_8/8_5_the_cochrane_collaborations_tool_for_assessing_risk_of_bias.htm)) y de EINA ([www.riskofbias.info](http://www.riskofbias.info)). Cualquiera sea el instrumento utilizado por los autores de la revisión, los evaluadores deberían sentirse satisfechos de que aborda los elementos enumerados en el ítem 9 del instrumento. Debe tenerse en cuenta que la guía que se proporciona aquí no es exhaustiva: AMSTAR-2 solo aborda los dominios de sesgo reconocidos con mayor frecuencia. Al revisar el riesgo de sesgo evaluado por los autores de la revisión, debe buscarse el consejo de expertos en métodos y de contenido, si ellos no están incluidos en su equipo, quienes deberían proporcionar consejos específicos sobre qué factores de confusión son importantes, cómo identificar los sesgos de selección y de medición que probablemente sean relevantes para la revisión y qué técnicas de ajuste de confusión serían apropiadas. La siguiente lista de dominios de riesgo de sesgo ha sido seleccionada del instrumento ROBINS-I / ACROBAT-NRSI, el más pertinente para RS que incluyen EINA:

**Confusión.** La confusión ocurre cuando los efectos de dos intervenciones asociadas o dos exposiciones (por ejemplo, el tabaquismo y el consumo de alcohol) no se han separado durante el análisis. Esto puede dar como resultado un efecto atribuido a una variable, cuando en realidad se debe a la otra. En los estudios de intervenciones, la confusión también puede estar relacionada con la indicación de tratamiento, por ejemplo, cuando un medicamento se administra preferentemente a pacientes con tasas más altas de comorbilidades que el fármaco de comparación y donde estas condiciones comórbidas están asociadas con el resultado de interés. Estas evaluaciones

usualmente se cuantifican con los datos de línea de base reportados en los estudios individuales. La posible confusión de línea base puede ser abordada de varias maneras, incluido el diseño (por ejemplo, al igualar por puntaje de propensión), el ajuste (por ejemplo, mediante regresión logística) y otras técnicas como el análisis de variables instrumentales y la inclusión de exposiciones 'trazadoras'. Es común suponer que, incluso con técnicas sofisticadas, no se ajustarán completamente todos los factores de confusión, lo que significa que son asociaciones débiles, incluso si son estadísticamente significativas, y deben interpretarse con cautela.

**Sesgo de selección de muestra.** Esto ocurre cuando los sujetos son muestreados de una manera sesgada que distorsiona directamente la verdadera relación entre la exposición y el resultado. No requiere un tercer factor, como en el caso de la confusión. Por ejemplo, no se debería estudiar la asociación entre el tabaquismo y las enfermedades del corazón al reclutar sujetos derivados a una clínica de cesación tabáquica. La selección de sujetos con la exposición de interés no debe estar relacionada con su resultado. Del mismo modo, la selección de sujetos con el resultado no debe estar relacionado con su estado de la exposición. El momento de la selección también puede ser importante. Si los sujetos han estado usando un medicamento durante algún tiempo antes del enrolamiento (usuarios frecuentes), serán un grupo tolerante con un menor riesgo de resultados adversos. Por esta razón, los estudios fármaco-epidemiológicos actuales reclutan 'nuevos usuarios' de medicamentos (análogo al inicio del tratamiento en un ECA). Otros sesgos de muestreo temporal (como el sesgo de tiempo inmortal y el sesgo de inicio) son a veces importantes. Se recomienda que los usuarios se dirijan al documento de orientación del ROBINS-I para obtener más información.

**Sesgo en la medición de exposiciones y resultados.** La medición de una exposición o del tratamiento puede clasificarse erróneamente si no se realiza un registro preciso en tiempo real. Por lo general, los estudios fármaco-epidemiológicos modernos usan registros de prescripciones o dispensación como sustituto de consumo. Pero la adherencia al tratamiento dispensado no será del 100%, por lo que el consumo real se clasificará erróneamente por este método; si el error no es diferencial, se subestimaría una eventual diferencia. Sin embargo, algunas preguntas de investigación dependen del recuerdo (p. ej., exposición ultravioleta y melanoma). Esto puede llevar a una clasificación errónea diferencial. Por ejemplo, los padres de un niño que ha muerto de síndrome de muerte súbita del lactante pueden tener mayor recuerdo de cualquier medicamento que administraron al bebé antes del evento.

**La medición de los resultados también puede verse afectada por clasificación errónea, que si fuera no diferencial subestimaría una eventual diferencia. Sin embargo, la clasificación errónea no diferencial puede introducir sesgos. Por ejemplo, si se realizan ecografías de miembros inferiores con más frecuencia en mujeres con piernas hinchadas y doloridas que están tomando anticonceptivos orales, la selección de individuos de una clínica de ecografías puede sesgar la asociación entre los anticonceptivos orales y la trombosis venosa profunda.**

**Reporte selectivo de resultados y análisis.** Los grandes estudios observacionales pueden analizar bases de datos poblacionales que registran múltiples resultados en una población definida. Si los resultados no están pre-especificados (preferiblemente en un protocolo registrado) los investigadores pueden tentarse de analizar múltiples resultados e informar selectivamente aquellos que parecen ser diferentes entre individuos expuestos y no expuestos. Además, habitualmente puede disponerse de varios métodos potenciales para analizar un conjunto de datos no aleatorizados (por ejemplo, diferentes formas de categorizar la intervención, o diferentes modelos de ajuste multivariable). Si el plan de análisis en el protocolo no se



especifica antes del estudio, puede ser posible seleccionar un conjunto de análisis que parezcan mostrar diferencias estadísticamente significativas no aparentes en los otros análisis. Los revisores deben determinar si los autores de los estudios pre-especificaron los resultados y los análisis. Esto será más fácil a medida que más estudios se registren antes de llevarse a cabo.

#### **10. ¿Los autores de la revisión reportaron las fuentes de financiación de los estudios incluidos en la revisión?**

Varias investigaciones han demostrado que es más probable que los estudios patrocinados comercialmente tengan hallazgos que favorecen el producto de un patrocinador que los estudios financiados de forma independiente. Es valioso que los autores de la revisión documenten las fuentes de financiamiento para cada estudio incluido en la revisión, o que registren que la información no se proporcionó en los reportes de los estudios. Dependiendo de esta información es posible analizar por separado los resultados de los estudios con financiación comercial de los estudios independientes.

#### **11. Si se realizó un meta-análisis, ¿los autores de la revisión usaron métodos apropiados para la combinación estadística de resultados? (Sólo se debe completar ese dominio si se reportó meta-análisis u otras técnicas de síntesis)**

Los autores de la revisión deberían haber declarado explícitamente en el protocolo de revisión los principios sobre los cuales basarían su decisión de realizar el meta-análisis de los datos de los estudios incluidos. Estos incluyen el deseo de obtener un único efecto combinado (por ejemplo, de una cantidad de estudios compatibles, pero con poco poder en forma individual) y la medida en que los estudios se considerarán compatibles (en términos de poblaciones, controles e intervención) y, por lo tanto, aptos para ser combinados. Además, cuando se consideró que era apropiado el meta-análisis, los autores deberían haber explicado sus decisiones sobre el uso de modelos de efectos fijos o aleatorios en el caso de los ECA, y establecer los métodos que pretendían usar para investigar la heterogeneidad. Con las poblaciones de estudio de los EINA, el tamaño varía desde pequeñas cohortes (con decenas o cientos de participantes) a estudios con cientos de miles de personas y miles de eventos. Si estos resultados se combinaran con los de los ECA más pequeños, las estimaciones de efecto estarían dominadas por los datos de los EINA.

Además, los resultados de los EINA pueden verse afectados por un amplio rango de sesgos (ver arriba), lo que significa que las estimaciones combinadas globales pueden ser precisas pero sesgadas. Los autores de la revisión deben informar las estimaciones combinadas por separado para los diferentes tipos de estudios. En el caso de los EINA, la combinación puede resultar en una estimación del efecto muy precisa y 'estadísticamente significativa', pero sesgada. Sin embargo, el intervalo de confianza se calcula sobre la asunción de que no hay sesgo (es decir, como si se obtuvieran de un ECA de alta calidad con el mismo número de participantes). Es raro que un EINA tenga tan bajo riesgo de sesgo como un ECA de alta calidad para la misma pregunta de investigación, por lo que, los intervalos de confianza para EINA (y las estimaciones combinadas basadas en EINA) deben considerarse con precaución. Este problema es importante al considerar la incertidumbre sobre el riesgo de sesgo, que es variable, entre los EINA.

La heterogeneidad es un problema importante en cualquier meta-análisis. Es particularmente importante en una RS de EINA, debido a los métodos más diversos que es probable que se hayan utilizado en los estudios. Además de las fuentes habituales de heterogeneidad (diferentes comparadores, variaciones en el riesgo basal de los resultados u otras características de la población del estudio,

diferentes intervenciones, como efectos de dosis, contexto/ámbito, experiencia de los médicos, y diferentes definiciones de resultados), es importante considerar otros factores como la heterogeneidad en el origen de los participantes, la integridad de los datos, los métodos de gestión de datos y el análisis.

El ajuste estadístico de los efectos de la intervención para los factores de confusión puede dar como resultado estimaciones bastante diferentes de la estimación no ajustada derivada de los datos sin procesar.

En general, cuando se combinan los resultados de los autores de la revisión de EINA, se debe combinar las estimaciones ajustadas de efecto, no los datos crudos, y en el caso de hacerlo, debería haber una clara justificación. Sin embargo, es probable que diferentes estudios informen los efectos del tratamiento ajustados para diferentes conjuntos de covariables (o covariables medidas o ajustadas de diferentes formas); esta diversidad representa otra fuente potencial de heterogeneidad.

#### **12. Si se realizó un meta-análisis, ¿los autores de la revisión evaluaron el impacto potencial del riesgo de sesgo en estudios individuales sobre los resultados del meta-análisis u otra síntesis de evidencia?**

En los casos en que los autores de la revisión hayan optado por incluir solo ECA de alta calidad, puede haber poca discusión sobre el impacto potencial de sesgo en los resultados. Pero si incluyeron ECA de calidad variable deben evaluar este impacto mediante el análisis de regresión o estimando tamaños de efectos combinando solo estudios con bajo riesgo de sesgo (análisis de sensibilidad). En el caso de considerar EINA, deben estimarse los tamaños de los efectos combinando solo estudios con bajo o bajo/moderado riesgo de sesgo. Si no se realizaron meta-análisis (u otras técnicas de síntesis de datos, como análisis de regresión) los autores aún deben proporcionar algunos comentarios sobre el probable impacto del riesgo de sesgo en los resultados de los estudios individuales.

#### **13. ¿Los autores de la revisión consideraron el riesgo de sesgo de los estudios individuales al interpretar / discutir los resultados de la revisión?**

Incluso si no se llevaron a cabo meta-análisis, los autores de la revisión deberían incluir en la discusión el impacto del riesgo de sesgo en la interpretación de sus resultados. Esto siempre es importante, pero especialmente cuando las revisiones incluyen ECA con riesgo de sesgo variable, y en cualquier revisión que incluya EINA. Esta discusión no debe limitarse al impacto del riesgo de sesgo en las estimaciones combinadas (ver arriba), sino también para explicar las diferencias entre los resultados de los estudios individuales. Los autores deberían hacer una consideración explícita del potencial riesgo de sesgo si hacen recomendaciones que pueden tener un impacto en la atención clínica o en políticas de salud.

#### **14. ¿Los autores de la revisión proporcionaron una explicación satisfactoria y discutieron cualquier heterogeneidad observada en los resultados de la revisión?**

Existen muchas más causas potenciales de heterogeneidad en los resultados de los EINA que en los de los ECA. Muchos de los factores considerados en este instrumento incluyen diferentes diseños de estudio, diferentes métodos de análisis, diferentes poblaciones y diferentes intensidades de la(s) intervención(es) o dosificaciones en el caso de drogas. Tanto los elementos PICO como los dominios de sesgo enumerados en el ítem 9 deberían considerarse como importantes fuentes potenciales de heterogeneidad en los resultados. Los autores de la revisión deberían explorar estas posibilidades y discutir el impacto de la heterogeneidad observada en las conclusiones y en eventuales recomendaciones.

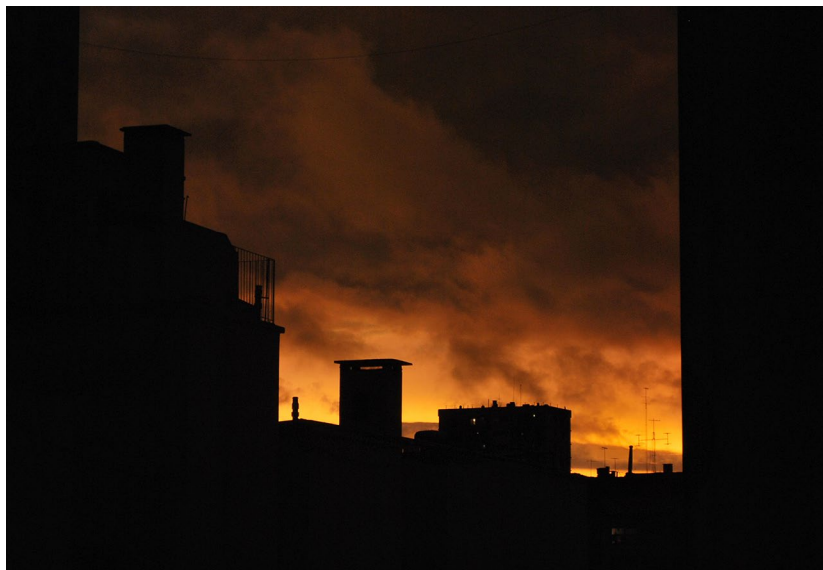
**15. Si se realizó una síntesis cuantitativa ¿los autores de la revisión llevaron a cabo una adecuada investigación de la posibilidad de sesgo de publicación (sesgo de estudio pequeño) y discutieron su impacto probable en los resultados de la revisión?**

Este es un tema muy importante, pero puede ser difícil para los autores de la revisión y para los evaluadores resolverlo de forma completa. Por lo general, se usan pruebas estadísticas o gráficos y, si estas pruebas son positivas, se considera que indican la presencia de sesgo de publicación. Sin embargo, las pruebas negativas no son una garantía de la ausencia de este tipo de sesgo. Hasta cierto punto, la importancia del sesgo de publicación depende del contexto y del ámbito.

Por ejemplo, una serie de estudios en apariencia metodológicamente sólidos, patrocinados por la industria (por ejemplo, estudios con drogas, dispositivos, o toxinas putativas) podrían ser más propensos a verse más afectados por el sesgo de publicación que estudios similares realizados sin financiamiento dependiente de la industria. Los aspectos clave incluyen: si los autores han hecho su mejor esfuerzo para identificar el sesgo de publicación a través de búsquedas bibliográficas más profundas e intensas (según sea necesario y de acuerdo con el ámbito), si mostraron el impacto probable del sesgo de publicación en su interpretación y discusión de los resultados, y si realizaron un análisis de sensibilidad para determinar cuántos estudios 'nulos' faltantes serían necesarios para invalidar los resultados obtenidos.

**16. ¿Los autores de la revisión informaron cualquier fuente potencial de conflicto de intereses, incluyendo cualquier financiamiento recibido para llevar a cabo la revisión?**

Como se señaló al comentar riesgo de sesgo, los estudios individuales financiados por intereses privados pueden generar resultados más propensos a favorecer la intervención que los estudios independientes. La misma suposición se aplica a las RS. Los autores deben reportar sus fuentes de financiamiento directo, y las revistas generalmente lo requieren para su publicación, pero la evaluación de los conflictos de intereses de los revisores no se detiene ahí. Deben informar sus otros lazos. La revisión puede ser financiada de forma independiente, pero sus autores pueden tener vínculos con empresas que fabrican o comercializan los productos incluidos en la revisión sistemática. Los conflictos de intereses profesionales son poderosos, pero más difíciles de discernir ya que rara vez se reportan. Cuando los investigadores tienen un involucramiento profesional de larga data en un campo de investigación, se puede manifestar un conflicto de interés de este tipo cuando una revisión confronta con sus creencias más arraigadas. Los posibles conflictos de interés profesionales serán difíciles de evaluar, pero se pueden inferir si los revisores han publicado en forma extensa en el campo que se está revisando y sus estudios están incluidos en la RS. Si bien puede argumentarse que los efectos del conflicto de interés podrían manifestarse como fallas en otros dominios de sesgo, los desarrolladores del AMSTAR-2 creemos que este ítem siempre debe clasificarse por separado.



Cortesía de Silvia Spina