

Herramienta de Riesgo de Sesgo en Red (RoB NMA)

Dominio 1. Intervenciones y geometría de la red

1.1 ¿Fueron razonables todas las intervenciones y comparadores para la población objetivo?

- V:** si todos los estudios incluyeron participantes similares o excluyeron a personas no aptas para las intervenciones.
- PV:** si parece probable que los estudios incluyeran participantes similares, incluso si no se muestra explícitamente.
- PF:** si existen dudas de que algunos participantes no serían elegibles para ciertas intervenciones.
- F:** si está claro que algunos participantes no serían elegibles para ciertas intervenciones recibidas.
- Sin información** respecto de la suficiencia de los datos para emitir un juicio.

1.2 ¿Todas las intervenciones elegibles fueron incluidas en la red?

- V:** No se excluyeron intervenciones de manera inapropiada.
- PV:** Se excluyeron intervenciones de manera apropiada y se proporcionó una justificación clara.
- PF:** Se excluyeron intervenciones sin una justificación clara que permita determinar si la exclusión fue apropiada.
- F:** Se excluyeron intervenciones de manera inapropiada, basándose en su probable impacto sobre los resultados del meta-análisis en red, con el fin de alinearlos con ideas preconcebidas sobre los efectos relativos de las intervenciones.
- Sin información:** Los datos reportados son insuficientes para emitir un juicio razonable.

1.3 ¿Fueron agrupadas apropiadamente las intervenciones en la red?

- V:** si la justificación para agrupar las intervenciones fue sólida y no se sugieren diferencias importantes en los efectos.
- PV:** si se combinaron intervenciones para las que son posibles diferentes efectos, pero es poco probable que ocurran.
- PF:** si se combinaron intervenciones que probablemente tengan efectos diferentes.
- F:** si se combinaron intervenciones que se sabe que tienen efectos diferentes.
- Sin información** si los datos son insuficientes para emitir un juicio.

1.4 ¿Estaban todas las intervenciones comparadas conectadas mediante comparaciones adecuadas dentro de los estudios?

- V:** si la red está conectada o si las comparaciones se hicieron dentro de una subred conectada.
- PV:** si se conectó la red agregando intervenciones para obtener evidencia indirecta.
- PF** si se tomaron medidas para conectar intervenciones desconectadas mediante datos observacionales, opinión de expertos u otros métodos.
- F:** si la red incluyó estudios de un solo brazo o se informaron resultados de comparaciones desconectadas.
- Sin información** si los datos son insuficientes para emitir un juicio.

Dominio 1: Bajo riesgo

Preocupaciones

Alto riesgo

Textuales:

Justificación del juicio:

Dominio 2: Modificadores de efecto

2.1 ¿Fueron similares las definiciones de desenlace y los puntos temporales entre comparaciones directas en la red?

- V:** si las definiciones de desenlace o los puntos temporales fueron idénticos o lo suficientemente similares como para no afectar los efectos de la intervención.
- PV:** si se usaron desenlaces o tiempos diferentes, pero es poco probable que afecten los efectos de la intervención.
- PF:** si se usaron desenlaces o tiempos diferentes que probablemente resulten en efectos distintos.
- F:** si se sabe que los desenlaces o tiempos diferentes provocan efectos diferentes.
- Sin información** si los datos son insuficientes para emitir un juicio.

2.2 ¿Fueron similares las características de los participantes que modifican el efecto en las comparaciones directas de la red?

- V:** si los estudios que realizan comparaciones diferentes fueron similares en las características conocidas o sospechadas que modifican el efecto.
- PV:** si los estudios fueron razonablemente similares en estas características.
- PF:** si se sospecha que las características no son razonablemente similares entre comparaciones.
- F:** si hay evidencia de variación importante entre las comparaciones en dichas características.
- Sin información** si los datos son insuficientes para emitir un juicio.

2.3 ¿Fueron similares las características de los estudios que modifican el efecto en las comparaciones directas de la red?

- V** si los estudios fueron similares en las características conocidas o sospechadas que modifican el efecto.
- PV:** si los estudios fueron razonablemente similares en la mayoría de estas características.
- PF:** si se sospecha que no fueron razonablemente similares.
- F:** si hay evidencia de variaciones importantes en dichas características entre las comparaciones.
- Sin información** si los datos son insuficientes para emitir un juicio.

2.4 Si la respuesta fue F o PF en 2.1, 2.2 o 2.3: ¿El análisis abordó adecuadamente las diferencias en los modificadores del efecto en la red?

- V:** si los autores utilizaron métodos aceptables para realizar el análisis, tales como: (i) **métodos estadísticos** (metarregresión, análisis de subgrupos o modelos de efectos mixtos) que probablemente hayan tenido en cuenta estas diferencias; o (ii) **análisis de sensibilidad** que demostraron que las diferencias eran poco relevantes en la revisión, ajustando por el conjunto completo de modificadores del efecto en cuestión y abordando la comparabilidad de los resultados con los análisis no ajustados.
- PV:** si los análisis fueron parcialmente explorados, pero quedan diferencias que no fueron abordadas o si los hallazgos de los análisis no fueron considerados adecuadamente.
- F / PF:** si no se exploraron análisis para abordar las diferencias en los posibles modificadores del efecto.
- Sin información:** cuando los datos disponibles son insuficientes para emitir un juicio razonable.

Dominio 2: Bajo riesgo Preocupaciones

Alto riesgo

Textuales:

Justificación del juicio:

Dominio 3: Síntesis estadística

3.1 ¿La síntesis respetó la aleatorización dentro de los estudios?

- V:** si se analizaron comparaciones dentro del mismo estudio o si se usaron datos a nivel de la rama o del paciente con un modelo de efecto fijo por estudio.
- PV:** si se usaron datos a nivel de la rama o del paciente con un modelo de efecto aleatorio y los resultados promedio fueron similares entre comparaciones.
- PF:** si se usaron modelos de efecto aleatorio y los resultados promedio fueron diferentes entre las comparaciones.
- F:** si los datos fueron combinados entre estudios antes de compararse entre intervenciones.
- Sin información:** si los datos son insuficientes para emitir un juicio.

3.2 ¿No se sospecha sesgo de publicación u otros sesgos por omisión selectiva de resultados?

- V:** La justificación para agrupar las intervenciones fue sólida, el enfoque fue pre especificado y no hay indicios de diferencias importantes en los efectos de las intervenciones agrupadas.
- PV:** La cantidad de evidencia potencialmente excluida del meta-análisis en red tendría un impacto trivial en los hallazgos.
- PF:** Faltan resultados elegibles en el meta-análisis en red y la cantidad de evidencia faltante podría afectar las estimaciones del efecto de las intervenciones.
- F:** Faltan resultados elegibles en el meta-análisis en red y es probable que estos resultados sean sistemáticamente diferentes de los incluidos, con suficiente evidencia faltante como para afectar las estimaciones del efecto de las intervenciones.
- Sin información:** No se reportan datos suficientes para emitir un juicio razonable.

3.3 ¿Sólo se reportaron todos los análisis predefinidos o se explicaron las discrepancias?

- V:** Se siguieron los análisis predefinidos conforme a un protocolo detallado y pree especificado.
- PV:** Hay indicios de que se siguieron los análisis predefinidos, o todas las desviaciones respecto de los métodos planificados fueron justificadas con razones no relacionadas con los resultados observados.
- PF:** No hay indicios de que se hayan seguido los análisis predefinidos, o no hay suficiente detalle sobre los métodos que probablemente fueron planificados e implementados.
- F:** No se siguieron los análisis predefinidos, o se utilizaron otros análisis, y las desviaciones respecto de los análisis planificados se hicieron en función de los resultados.
- Sin información:** No se reportan datos suficientes para emitir un juicio razonable.

3.4 ¿Los sesgos en los estudios primarios fueron mínimos o se abordaron en la síntesis?

- V:** Todos los estudios fueron evaluados como de bajo riesgo de sesgo, o los análisis de sensibilidad demostraron que incluir estudios con mayor riesgo no tuvo impacto en los resultados.
- PV:** La proporción de información con alto riesgo de sesgo fue demasiado pequeña como para afectar los resultados, o los análisis de sensibilidad o de umbral demostraron un impacto mínimo al incluir estos estudios.
- PF:** La proporción de información con alto riesgo de sesgo fue suficiente como para afectar los resultados, y los análisis de sensibilidad no demostraron un impacto mínimo, o no se evaluó el riesgo de sesgo de los estudios incluidos.
- F:** Existen sesgos importantes en los estudios primarios que no fueron abordados por los revisores, o los análisis de sensibilidad muestran que los resultados están fuertemente influenciados por estudios con alto riesgo de sesgo.
- Sin información:** No se reportan datos suficientes para emitir un juicio razonable.

3.5 ¿Se usaron métodos apropiados para manejar estudios con múltiples brazos?

- V:** No se incluyeron estudios con múltiples brazos, o los datos de estos estudios fueron adecuadamente considerados en el análisis.
- PV:** El software, código o métodos del meta-análisis en red no fueron reportados en detalle, pero otras evidencias sugieren que se utilizaron herramientas y métodos apropiados.
- PF:** Los estudios con múltiples brazos no fueron representados adecuadamente, o los métodos elegidos no consideraron correctamente las correlaciones en los efectos aleatorios.
- F:** Los estudios con múltiples brazos no fueron representados adecuadamente, o los métodos utilizados no indicaron un manejo apropiado de las correlaciones.
- Sin información:** No se reportan datos suficientes para emitir un juicio razonable.

3.6 ¿Se asumieron métodos apropiados sobre la homogeneidad o heterogeneidad de los efectos dentro de las comparaciones?

- V o PV:** Si no se modeló la heterogeneidad (por ejemplo, mediante un modelo de efectos fijos para eventos raros) y esto estuvo bien justificado, o si se modeló la heterogeneidad no explicada (es decir, se usó un modelo de efectos aleatorios) y no hay evidencia en los datos que contradiga los supuestos del modelo. También se considera apropiado si se realizó un análisis de sensibilidad con un enfoque más general o adecuado y los resultados fueron similares a los del análisis principal.
- F o PF:** Si no se modeló la heterogeneidad y esto no fue bien justificado, o si se modeló la heterogeneidad y hay evidencia en los datos que contradice los supuestos del modelo.
- Sin información:** No se reportan datos suficientes para emitir un juicio razonable.

3.7 ¿No hubo evidencia de conflicto entre las estimaciones directas e indirectas de un mismo efecto

- V:** No existían fuentes potenciales de inconsistencia en la red (por ejemplo, no había bucles cerrados, o estos estaban formados solo por comparaciones dentro de un mismo estudio con múltiples brazos), o se exploró adecuadamente la inconsistencia y no hubo indicios de su presencia. Esto normalmente requiere una base de evidencia suficientemente amplia para superar los problemas de baja potencia en los tests de inconsistencia.
- PV:** Se exploró adecuadamente la inconsistencia y no se encontró evidencia de su existencia. Esto puede incluir casos donde algunas estimaciones sean extremas (por ejemplo, por celdas con valor cero), pero la dirección general del efecto sea coherente.
- PF:** Hubo algo de evidencia de inconsistencia, aunque no pareció grave, o existen preocupaciones relevantes sobre la baja potencia de los tests utilizados para detectar inconsistencia.
- F:** No se brindó información sobre la inconsistencia o existe evidencia clara de su presencia.
- Sin información:** No se reportan datos suficientes para emitir un juicio razonable.

3.8 Si hubo conflicto entre evidencia directa e indirecta: ¿se abordaron adecuadamente los resultados contradictorios?

- V o PV:** Si los análisis alternativos tuvieron una justificación clara y resolvieron el conflicto original.
- F o PF:** Si se compararon los resultados del meta-análisis directo con los del meta-análisis en red o con la literatura previa (por ejemplo, revisiones sistemáticas), o si se excluyeron aleatoriamente estudios primarios hasta lograr consistencia.
- Sin información:** No se reportan datos suficientes para emitir un juicio razonable.

3.9 Si se realizó un análisis bayesiano: ¿fue apropiada la elección de las distribuciones a priori?

- No aplicable:** Si no se realizó un análisis bayesiano (es decir, si se usó un enfoque frecuentista).
- V:** Si se utilizaron distribuciones a priori claramente no informativas en el rango de parámetros posibles, o si se usaron distribuciones informativas con justificación clara, o si se realizó un análisis de sensibilidad con distribuciones a priori alternativas adecuadas y los resultados fueron similares a los principales.
- PV:** Si se usaron distribuciones a priori no informativas estándar.
- F o PF:** Si se usaron distribuciones a priori informativas sin una justificación clara.
- Sin información:** No se reportan datos suficientes para emitir un juicio razonable.

Dominio 3: Bajo riesgo

Preocupaciones

Alto riesgo

Textuales:

Justificación del juicio:

Riesgo de sesgo general en el meta-análisis en red

Sesgo en los resultados del meta-análisis en red

- Alto riesgo de sesgo:** Si **cualquier dominio**, ya sea del instrumento ROBIS/AMSTAR 2 o del instrumento RoB NMA, fue juzgado con **alto riesgo de sesgo**.
- Bajo riesgo de sesgo:** Si **todos los dominios**, tanto del instrumento ROBIS/AMSTAR 2 como del RoB NMA, fueron juzgados con **bajo riesgo de sesgo**.
- Algunas preocupaciones:** Si **al menos un dominio** de cualquiera de los dos instrumentos fue juzgado como "algunas preocupaciones" o "alto riesgo de sesgo".

Sesgo en las conclusiones del meta-análisis en red

Los evaluadores también pueden emitir un juicio general sobre el sesgo en las conclusiones de los autores, evaluando la interpretación de los hallazgos. Esto incluye examinar si se abordaron las preocupaciones identificadas, si se consideró la relevancia de los estudios (como se recomienda en la Fase 1 de ROBIS), y si los autores enfatizaron resultados únicamente por su significancia estadística.

Los ítems 3.5, 3.6 y 3.9 deben ser reconsiderados al evaluar el riesgo de sesgo en las conclusiones, ya que decisiones inadecuadas en el modelado estadístico pueden llevar a subestimar la incertidumbre en los resultados, lo que probablemente genere sesgo en las conclusiones del NMA.

- Preocupaciones
 - Sin preocupaciones
-

Guía para utilizar la herramienta RoB NMA

Es recomendable que el equipo evaluador tenga experiencia en la realización de MAR y en los sesgos inherentes a estos, así como experiencia clínica en la condición en estudio para evaluar las preguntas de señalización relacionadas con los modificadores de efecto. Cualquier protocolo o constancia de documentación en PROSPERO (Registro Prospectivo Internacional de Revisiones Sistemáticas) debe obtenerse y leerse junto con la revisión sistemática con MAR.

Evaluar el sesgo en un MAR cuando la información no se incluye en el manuscrito puede ser difícil. A menudo, la información faltante puede obtenerse de archivos complementarios o del registro del estudio. También se puede contactar a los autores para solicitar información adicional o aclaraciones sobre sus métodos. Se puede asignar la

calificación "sin información" si no se dispone de información sobre los métodos utilizados para un elemento específico de la herramienta.

Dependiendo del objetivo de la evaluación, detenerla puede ser razonable si se encuentra un alto riesgo de sesgo en un dominio en particular, reduciendo así el esfuerzo necesario para llegar a un juicio sobre el riesgo de sesgo. En algunos casos, puede ser posible comenzar con afirmaciones particulares que se espera que identifiquen un alto riesgo de sesgo. Por ejemplo, si se ha utilizado un enfoque incorrecto para el análisis, como agregar los resultados de los participantes en los ensayos por separado para cada grupo de intervención (como si los datos provinieran de un único ensayo grande y simple en lugar de múltiples ensayos), entonces la afirmación de señalización 3.1 puede responderse como falsa (es decir, no se preservó la aleatorización dentro del estudio). Por lo tanto, la evaluación RoB NMA puede detenerse porque los resultados del MAR podrían juzgarse con alto riesgo de sesgo. En tal situación, es probable que las conclusiones del MAR también presenten un alto riesgo de sesgo.

Dominio 1: intervenciones y geometría de red

El primer dominio analiza cómo se seleccionan y agrupan las intervenciones y si constituyen un conjunto apropiado de intervenciones para realizar una NMA.

1.1 Todas las intervenciones y sus comparadores incluidos en el NMA son alternativas razonables para toda la población objetivo

Para garantizar la comparabilidad de los efectos relativos en toda la red, todos los participantes incluidos deben ser elegibles para recibir todas las intervenciones al inicio del estudio (incluyendo tratamiento activo y placebo, ausencia de tratamiento, control, atención habitual o comparador). Esta idea se conoce a veces como aleatorización conjunta. Por ejemplo, la cirugía, la quimioterapia y el tratamiento para el control de síntomas son tratamientos para el cáncer de pulmón de células no pequeñas, pero estos tratamientos no se considerarían aleatorizables conjuntamente porque son adecuados para diferentes grupos de pacientes: la quimioterapia suele considerarse después de la cirugía, y el control de síntomas es una estrategia para la enfermedad en estadio avanzado.

Un MAR podría considerarse como un intento de imitar un gran ensayo controlado aleatorizado (ECA) con un único conjunto de criterios de elegibilidad para la inscripción, comparando todas las intervenciones de interés. Cuanto más se asemejen los participantes y las intervenciones del MAR a un ensayo hipotético de este tipo, menor será el riesgo de sesgo.

Fuente de sesgo

Si diferentes intervenciones en la red son adecuadas para diferentes tipos de personas (por ejemplo, algunas intervenciones son para tratamiento de primera línea y otras para tratamiento de segunda línea), lo más probable es que los participantes difieran en las diferentes comparaciones de intervenciones, lo que lleva al fracaso del supuesto de transitividad y al sesgo posterior.

¿Cómo evaluar esta afirmación?

Algunas intervenciones presentan contraindicaciones (es decir, la intervención es inapropiada para ciertos tipos de pacientes debido a su etapa de la enfermedad, antecedentes de tratamiento, efectos adversos o debido a una menor efectividad esperada). Los autores de la revisión deben indicar las contraindicaciones y demostrar que las personas contraindicadas están excluidas de todos los estudios. Se requiere experiencia en el tema para evaluar esta información.

Por ejemplo, una revisión de intervenciones terapéuticas para personas con insuficiencia de convergencia (una afección ocular que afecta la forma en que los ojos trabajan juntos al mirar objetos cercanos) incluyó todas las intervenciones no quirúrgicas: ejercicios de convergencia en el hogar (flexiones con lápiz o diana), tratamiento de convergencia o acomodación por computadora en el hogar, tratamiento de convergencia o acomodación en el consultorio, anteojos de lectura con prisma de base hacia adentro, anteojos con prisma de base hacia adentro combinados con lentes de adición progresiva e intervención placebo o simulada. Estas intervenciones están

disponibles para todos los pacientes con un diagnóstico de la afección y, por lo tanto, los pacientes se consideraron aleatorizables de forma conjunta.

Responda **verdadero (V)** si los autores de la revisión proporcionan evidencia de que todos los estudios incluyen participantes similares o que los individuos que poseen contraindicaciones para recibir cierta intervención fueron excluidos de todos los estudios.

La respuesta es probablemente verdadera (**PV**) si parece probable desde el punto de vista clínico que los estudios hayan incluido participantes similares o que los individuos contraindicados para una intervención hayan sido excluidos de todos los estudios, aunque los autores no estén mostrando esta información explícitamente.

La respuesta probablemente sea verdadera (**PF**) si existen preocupaciones de que algunos participantes no sean elegibles para algunas de las intervenciones.

Responda **falso (F)** si está claro que algunos participantes no serían elegibles para algunas de las intervenciones.

Responda "sin información" (**SI**) cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

1.2 Todas las intervenciones elegibles fueron incluidas en la red

La investigación empírica ha demostrado que la exclusión de una intervención elegible (ya sea una intervención activa o un comparador) de una red puede tener un efecto importante en las estimaciones de los efectos de la intervención en un MAR. La principal preocupación aquí es que las intervenciones pueden excluirse en función de su efecto previsto o conocido sobre los resultados del MAR, para garantizar que los resultados concuerden con las opiniones que tiene a priori el grupo investigador sobre los efectos relativos de las intervenciones.

Fuente de sesgo

La exclusión de intervenciones del MAR puede deberse a descuidos de los autores o a razones pragmáticas (p. ej., limitaciones de recursos). Es improbable que el primer caso introduzca sesgo, aunque podría limitar la utilidad del MAR. El segundo caso puede o no introducir sesgo, dependiendo de si la elección de las intervenciones a excluir se relaciona con los resultados esperados de haberlas incluido. Dependiendo de la estructura de la red, las exclusiones podrían afectar sustancialmente los efectos estimados de las intervenciones y pueden afectar la clasificación de las intervenciones según su efectividad o seguridad.

¿Cómo evaluar esta afirmación?

Para garantizar la transparencia y la integridad de los informes, los autores del MAR deben aclarar qué intervenciones son elegibles para su inclusión en el MAR y la justificación de esta elección. La comparación de las intervenciones que habían planificado incluir (es decir, como se encuentran en el protocolo de revisión) con las incluidas en la red podría revelar inconsistencias, y dichas inconsistencias deben investigarse. Los autores del MAR pueden proporcionar una lista de cualquier estudio que tenga las intervenciones elegibles que fueron excluidas del MAR, con una justificación para cada decisión tomada. Examinar la justificación para la exclusión de las intervenciones elegibles de la red es importante para determinar si dichas decisiones están adecuadamente justificadas. Las exclusiones pueden ser razonables (p. ej., si surgen fuertes preocupaciones sobre la violación de los supuestos del MAR).

V: si no se excluyeron intervenciones de manera inapropiada.

PV: si se excluyeron las intervenciones de manera apropiada y se dio una justificación clara.

PF: si se excluyeron intervenciones sin una justificación clara para determinar si dicha acción fue apropiada o no.

F: si las intervenciones se excluyeron de manera inapropiada (es decir, se excluyeron en función de su probable efecto sobre los resultados del MAR, para garantizar que los resultados coincidieran con las opiniones a priori sobre los efectos relativos de las intervenciones).

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

1.3 Las intervenciones se agruparon adecuadamente en nodos de la red

Es inusual que dos estudios diferentes implementen la misma intervención activa de forma idéntica, al igual que la implementación en dos centros distintos dentro de un estudio probablemente difiera. Una posible excepción podrían ser las intervenciones farmacéuticas que examinan el mismo fármaco en la misma formulación, la misma dosis y la misma pauta de administración. Por lo tanto, para un MAR suele ser necesario agrupar las intervenciones (o sus implementaciones específicas) en nodos. Este requisito se aplica de forma similar a la agrupación de cualquier intervención de control incluida en la red, incluyendo placebos, ausencia de tratamiento, control o atención habitual. La agrupación de intervenciones o sus variantes se denomina a veces agrupación (“lumping” del inglés). Los autores de MAR pueden utilizar tres enfoques posibles para agrupar diferentes intervenciones o variantes en un MAR: (i) un enfoque de agrupación amplio que agrupa intervenciones similares a un nivel amplio y resulta útil para estimar los efectos de clase de una intervención; (ii) un enfoque de agrupación y desmantelamiento de componentes, basado en metarregresión, para investigar los efectos atribuidos a diferentes componentes; y (iii) un enfoque de modelado jerárquico (efecto de clase) que agrupa intervenciones similares como una clase, pero asume variaciones en los efectos entre estas intervenciones, utilizando el modelado para estimar los efectos de intervenciones específicas dentro de las clases.

Fuente de sesgo

Al agrupar las intervenciones, estas deben ser suficientemente similares (es decir, no se espera que difieran en sus efectos relativos). Podría surgir un sesgo en la pregunta objetivo si el efecto estimado por las intervenciones combinadas no refleja el efecto que se observaría si se hubiera utilizado la intervención objetivo en forma separada (no combinada).

¿Cómo evaluar esta afirmación?

La evidencia de la similitud de los efectos de las intervenciones para justificar la agrupación de intervenciones puede provenir de investigaciones de heterogeneidad en meta-análisis por pares o incluso en estudios individuales. Por ejemplo, los análisis dosis-respuesta pueden indicar que diferentes dosis de una intervención son equivalentes, lo que justifica su combinación en un nodo de red. Las estimaciones del efecto de la intervención pueden ser insesgadas siempre que la relación dosis-respuesta (o la evolución temporal) sea especificada en forma correcta.

A menudo, la evidencia empírica que sustenta la agrupación de intervenciones basándose explícitamente en la similitud de efectos es insuficiente. Por lo tanto, puede requerirse una experiencia clínica adecuada para emitir el juicio. También debe considerarse cualquier dato cuantitativo presentado por los autores que respalde la pertinencia de la agrupación de intervenciones. La posible agrupación de intervenciones puede ser beneficiosa cuando estas decisiones son debatidas por expertos clínicos y metodológicos del equipo de investigación, y lo ideal es que hayan sido especificadas previamente para reducir el uso indebido del análisis de datos para encontrar patrones.

V: si la justificación para agrupar las intervenciones fue sólida, el enfoque fue especificado previamente y no hubo indicios evidentes de diferencias importantes en los efectos de las intervenciones agrupadas, o si no hubo agrupación de diferentes intervenciones.

PV: si se han combinado intervenciones para las cuales son posibles efectos diferentes, pero poco probables.

PF: si se han combinado intervenciones que probablemente tengan efectos diferentes.

F: si se han combinado intervenciones que se sabe que tienen diferentes efectos.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

1.4 Todas las intervenciones comparadas se conectaron a través de una cadena adecuada de comparaciones dentro del estudio

Las redes de intervenciones pueden estar desconectadas o débilmente conectadas por diversas razones. Por ejemplo, es posible que no se hayan realizado estudios para comparar las intervenciones de interés. Incluso si se han realizado comparaciones, puede que no sea apropiado combinar los estudios en una sola red. En una red desconectada, algunas intervenciones no pueden compararse, porque no existe una cadena de comparaciones dentro del estudio que las una. Los esfuerzos por conectar una red que de otro modo estaría desconectada pueden

aumentar el conjunto de posibles comparaciones de intervenciones en una red, lo que permite extraer conclusiones sobre los efectos relativos de una gama más amplia de intervenciones. Sin embargo, los métodos deben seleccionarse con cuidado para evitar la introducción de sesgos en la síntesis.

Fuente de sesgo

Algunos métodos disponibles para conectar una red que de otro modo estaría desconectada introducen evidencia adicional con mayor riesgo de sesgo. Por ejemplo, se puede añadir evidencia no aleatoria a una red de ensayos controlados aleatorios, o se puede utilizar la opinión de expertos para informar una o más comparaciones.

¿Cómo evaluar esta afirmación?

Los autores pueden haber utilizado uno de varios enfoques para conectar redes desconectadas o débilmente conectadas descritas en la literatura. Una posibilidad es incluir intervenciones adicionales en la red que no sean de interés principal. Como alternativa, los autores pueden utilizar evidencia externa a los estudios incluidos a partir de la revisión principal, para cerrar las brechas en la red (o fortalecer los eslabones débiles). Por ejemplo, en una red de ECAs se puede incorporar evidencia observacional (p. ej., datos de registro) para comparaciones clave que carecen de evidencia de ECAs, se puede utilizar evidencia de otras poblaciones, o se puede obtener la opinión de expertos. Los autores de MAR pueden haber utilizado métodos de ajuste poblacional cuando surgen redes mal conectadas porque los estudios identificados habían examinado diferentes poblaciones.

Los autores del MAR también podrían haber abordado el problema mediante el modelo de análisis. Una opción es el modelado jerárquico (o modelos de clase), que generalmente asume que los grupos de intervenciones se comportan de manera similar; o bien el MAR de componentes, un enfoque especialmente adecuado para intervenciones complejas. Las secciones de métodos y resultados del MAR deben detallar si se encontraron redes inconexas durante los procesos de revisión y análisis, y cualquier intento de conectarlas. Si los autores conectaron la red agrupando las intervenciones incluidas, este enfoque debe evaluarse en el punto 1.3.

V: si está claro que la red está conectada o si la red tuvo algunas intervenciones desconectadas pero los autores solo hicieron comparaciones entre intervenciones en una subred conectada.

PV: si los autores conectaron la red agregando intervenciones adicionales con el propósito de crear evidencia indirecta sobre las principales comparaciones de interés.

PF: si los autores informan que se tomaron medidas para conectar intervenciones desconectadas con enfoques que implicaban el uso de datos observacionales, opinión de expertos, préstamo de datos de condiciones relacionadas u otros métodos similares.

F: si la red incluyó estudios de un solo brazo o si los resultados se informaron para comparaciones desconectadas.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

Los evaluadores deben decidir qué riesgo atribuir a la revisión (alto/bajo/algunas preocupaciones) con la cita textual del manuscrito y su justificación para su elección.

Dominio 2: modificadores de efecto

El segundo dominio se centra en los modificadores del efecto. En el contexto de un meta-análisis, un modificador del efecto es una característica a nivel de un estudio que causa heterogeneidad en los efectos de la intervención entre los estudios. En un MAR, el supuesto de transitividad se cumple cuando los modificadores del efecto se distribuyen de forma equivalente entre las comparaciones. Si bien los estudios incluidos en un meta-análisis inevitablemente diferirán en muchos aspectos, es probable que sólo un subconjunto de estas diferencias se deba a los modificadores del efecto. La variación en los modificadores del efecto debe mantenerse lo más baja posible para minimizar la heterogeneidad y garantizar la validez de cualquier meta-análisis (meta-análisis por pares o MAR).

En un análisis de modelo no lineal (MAR), los modificadores de efecto también pueden afectar la validez del supuesto de transitividad si existen diferencias sistemáticas en los modificadores de efecto entre las diferentes comparaciones

directas de la red. Los resultados de un MAR generalmente se consideran más sensibles a las diferencias en los modificadores de efecto entre comparaciones directas (a menudo denominadas incoherencia o inconsistencia) que a las diferencias dentro de las comparaciones directas (generalmente denominadas heterogeneidad).

Las tres primeras preguntas de señalización en este dominio buscan identificar si existen diferencias importantes en los modificadores del efecto al comparar tres tipos de modificadores por separado: definiciones de resultados y puntos temporales, características de los participantes y características del estudio. Los modificadores del efecto deben ser suficientemente similares en los estudios incluidos, incluso en un meta-análisis por pares. En este caso, "suficientemente similar" significa que son similares en aspectos que influirán en los efectos de las intervenciones. La cuarta afirmación se utiliza para determinar si los autores del MAR pudieron superar los problemas identificados en los tres primeros ítems. Esta afirmación sólo se considera si se identifica un problema en uno de los tres primeros ítems.

2.1 Las definiciones de resultados y los puntos temporales fueron similares en las comparaciones directas en la red

En un meta-análisis, suele esperarse cierto grado de variabilidad en las definiciones de resultados y los puntos temporales. La pregunta específica que responde esta afirmación es si existen diferencias sistemáticas en las definiciones de resultados, incluidos los métodos de medición y los puntos temporales (p. ej., respuesta a las seis semanas *frente* a los 12 meses) en las diferentes comparaciones directas realizadas por los estudios de la red. Las diferencias en las definiciones de resultados o los puntos temporales pueden abordarse en el MAR mediante metarregresión o ciertos tipos de modelos de efectos mixtos. Como alternativa, se pueden realizar análisis de subgrupos para analizar por separado las diferentes definiciones de resultados o los puntos temporales, o bien, el análisis de sensibilidad puede mostrar que el efecto de incluir diferentes definiciones de resultados y puntos temporales sería mínimo.

Fuente de sesgo

Si las definiciones de los resultados o los puntos temporales difieren sistemáticamente entre las comparaciones en la red de maneras asociadas con los efectos de la intervención, el supuesto de transitividad podría no ser válido. Combinar los resultados con un MAR produciría resultados incorrectos. Por ejemplo, si los estudios que comparan la intervención A con la intervención B miden principalmente la ansiedad, y los estudios de B versus C miden principalmente la depresión, una comparación indirecta de estos resultados para estimar el efecto de la intervención A versus C producirá una estimación sesgada tanto del efecto sobre la depresión como del efecto sobre la ansiedad si las intervenciones tienen diferentes efectos en estos dos resultados. Este hallazgo se debe a que el supuesto de transitividad no se cumple en todas las fuentes de evidencia directa y es un problema a pesar de la similitud de los resultados dentro de las comparaciones directas A versus B y A versus C.

¿Cómo evaluar esta afirmación?

La inspección de las definiciones de resultados y los puntos temporales utilizados en los estudios incluidos debe realizarse a nivel agregado de las comparaciones directas. La transitividad puede evaluarse comparando la distribución de los posibles modificadores del efecto de los resultados y los puntos temporales en las comparaciones directas disponibles, cuando se disponga de datos suficientes. Esta evaluación requiere criterio clínico. Se recomienda a los autores de MAR que limiten su análisis principal a un punto temporal y que realicen análisis de sensibilidad para mostrar cómo los resultados pueden verse afectados por puntos temporales alternativos.

V: si los estudios que realizaron diferentes comparaciones utilizaron definiciones de resultados o puntos temporales que eran idénticos o lo suficientemente similares como para que los efectos de la intervención no se vieran afectados por estas diferencias.

PV: si los estudios que hicieron diferentes comparaciones utilizaron diferentes resultados o puntos temporales y es poco probable que las diferencias sean consecutivas a diferencias en el efecto de la intervención.

PF: si los estudios que hicieron diferentes comparaciones de intervenciones utilizaron diferentes resultados o puntos temporales y es probable que estos den como resultado diferentes efectos de intervención.

F: si los estudios que realizan diferentes comparaciones de intervenciones utilizaron diferentes resultados o puntos temporales y se sabe que estos dan como resultado diferentes efectos de intervención.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

2.2 Las características de los participantes que modificaron el efecto fueron similares en las comparaciones directas en la red

Las características de los participantes en los estudios incluidos deben ser suficientemente similares en todos los estudios de cualquier meta-análisis. La pregunta específica que responde esta afirmación es si existen diferencias sistemáticas en las características de los participantes, conocidas o presuntamente modificadoras del efecto, en las diferentes comparaciones directas realizadas por los estudios de la red. Ejemplos de características de los participantes que podrían ser modificadores del efecto incluyen la edad, el sexo, el historial de enfermedades, la gravedad de estas y el historial de tratamiento. Sin embargo, se pueden realizar comparaciones indirectas válidas si los estudios difieren en características de los participantes que no son modificadores del efecto. Por ejemplo, si una intervención funciona de manera similar en hombres y mujeres, incluir estudios con diferentes proporciones de hombres y mujeres en diferentes comparaciones puede proporcionar estimaciones no sesgadas de los efectos relativos de la intervención.

Fuente de sesgo

Si las características de los participantes que modifican el efecto difieren sistemáticamente entre las comparaciones en la red, el supuesto de transitividad podría no ser válido. Combinar los datos con un MAR produciría resultados sesgados. Por ejemplo, si los estudios que comparan la intervención A con la intervención B incluyen principalmente participantes con enfermedad grave, y los estudios de B versus C incluyen principalmente participantes con enfermedad moderada, una comparación indirecta de estos resultados para estimar el efecto de la intervención de A versus C producirá una estimación sesgada del efecto tanto para los participantes con enfermedad grave como para los con enfermedad moderada si las intervenciones tienen efectos diferentes en estas dos poblaciones. Este hallazgo se debe a que el supuesto de transitividad no se cumple en todas las fuentes de evidencia directa.

¿Cómo evaluar esta afirmación?

El primer paso es identificar las características de los participantes que probablemente modifiquen el efecto (es decir, que afecten la eficacia y/o la seguridad de las intervenciones en la red). Este enfoque requiere juicio clínico y puede beneficiarse de la revisión de la literatura pertinente. Idealmente, los posibles modificadores del efecto se establecerán a priori en un protocolo para el MAR, incluyendo el método de identificación (p. ej., una revisión sistemática o la opinión de un experto clínico). La inspección de estas características de los participantes que modifican el efecto en los estudios incluidos debe realizarse a nivel agregado de las comparaciones directas para garantizar que el supuesto de transitividad sea plausible. Este enfoque también puede requerir juicio clínico. La información sobre los modificadores del efecto puede estar disponible en tablas que detallen los estudios incluidos o en una comparación descriptiva exhaustiva de las diferencias en las características de los pacientes.

V: si los estudios que realizaron diferentes comparaciones fueron similares en las características de los participantes que modificaron los efectos conocidos y sospechados.

PV: si los estudios que hicieron diferentes comparaciones fueron razonablemente similares (o se juzgan como razonablemente similares) en las características de los participantes que modificaron los efectos conocidos o sospechados.

PF: si surgen sospechas de que las características de los participantes que modificaron el efecto no fueron razonablemente similares en las comparaciones.

F: si existe evidencia de variación importante entre comparaciones en características de los participantes que modifican efectos conocidos o sospechados.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

2.3 Las características del estudio que podrían ser modificadoras del efecto fueron similares en las comparaciones directas en la red

Entre las características del estudio que podrían ser modificadoras del efecto se incluyen las características del diseño (como la aleatorización por conglomerados o un período de pre-inclusión), el entorno del estudio (p. ej., atención primaria o secundaria) o las condiciones del estudio (p. ej., verano o invierno en infecciones respiratorias). La pregunta específica que responde esta afirmación es si existen diferencias sistemáticas en las características del estudio que modifican el efecto entre las diferentes comparaciones directas realizadas por los estudios de la red.

Si los supuestos del MAR (es decir, transitividad o consistencia) no se cumplen, los hallazgos podrían estar sesgados. Los supuestos subyacentes al análisis estadístico proporcionan el marco para la validez de los resultados obtenidos. El incumplimiento de estos supuestos podría afectar la precisión o producir resultados incorrectos. Sin embargo, se pueden realizar comparaciones indirectas válidas si los estudios difieren en características que no son modificadoras del efecto. Por ejemplo, el uso de la agrupación en clústeres podría no afectar los efectos de la intervención. En este caso, no importa si los estudios A versus B utilizaron mayoritariamente la agrupación en clústeres y los estudios A versus C, no lo hicieron.

Fuente de sesgo

Las características de los estudios que modifican el efecto pueden variar entre estudios para una comparación particular (como con el meta-análisis por pares estándar, lo que causa heterogeneidad), pero también entre comparaciones (lo que causa inconsistencia) en la red. Si existe un desequilibrio en la distribución de los modificadores del efecto entre diferentes tipos de comparaciones directas, el MAR estará sesgado. Por ejemplo, si los estudios que comparan la intervención A con la intervención B fueron realizados principalmente en atención primaria y los estudios de B versus C fueron realizados principalmente en atención secundaria, una comparación indirecta de estos resultados para estimar el efecto de la intervención A versus C probablemente estará sesgada para cualquiera de estos entornos, si los efectos de la intervención tienden a diferir entre los entornos. Este hallazgo se debe a la falla de la transitividad (o consistencia) en todas las fuentes de evidencia directa.

¿Cómo evaluar esta afirmación?

El primer paso es identificar las características del estudio que probablemente modifiquen el efecto. Este enfoque probablemente requiera juicio clínico. Las más probables son características específicas del diseño asociadas con diferentes magnitudes de efecto. Idealmente, los posibles modificadores del efecto a nivel de estudio se establecerán a priori en un protocolo para el MAR (p. ej., una revisión sistemática independiente o la aportación de expertos).

La inspección de estas características de los estudios modificadores del efecto en los estudios incluidos debe realizarse a nivel agregado de las comparaciones directas para garantizar la verosimilitud de los supuestos de transitividad. La información sobre los modificadores del efecto puede estar disponible en las tablas que detallan los estudios incluidos o en una comparación descriptiva exhaustiva de las diferencias entre los estudios que realizan distintas comparaciones directas. Este punto no incluye el riesgo de sesgo en el estudio primario, que se aborda en la pregunta 3.3.

V: si los estudios que realizaron diferentes comparaciones directas fueron similares en las características de estudio modificadoras de efectos conocidos y sospechados (incluida la consideración de la naturaleza y dirección de cualquier sesgo).

PV: si los estudios que hicieron diferentes comparaciones directas fueron razonablemente similares (o se juzga que son razonablemente similares) en la mayoría de las características de estudio modificadoras de efectos conocidas o sospechadas.

PF: si surge la sospecha de que las características del estudio modificadoras del efecto no fueron razonablemente similares en las comparaciones directas.

F: si existe evidencia de variación importante en comparaciones directas en características del estudio modificadoras de efectos conocidos o sospechados.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

2.4 Si es F o PF para 2.1, 2.2 o 2.3: el análisis examinó adecuadamente las diferencias en los modificadores de efecto en toda la red.

Las diferencias en los modificadores de efecto en diferentes partes de la red pueden llevar a una falta de transitividad en la red, que puede no ser identificada por pruebas estándar de inconsistencia. Por el contrario, las diferencias sistemáticas en las características modificadoras de efecto en las diferentes comparaciones directas hechas por estudios en la red no resultan en sesgo si se tienen en cuenta en el análisis. Cuando una red no tiene muchos bucles cerrados, las pruebas estadísticas globales de inconsistencia pueden no ser informativas y se deben preferir las verificaciones de inconsistencia basadas en bucles, aunque generalmente tienen bajo poder estadístico para detectar inconsistencias. Cuando las pruebas estadísticas formales no son concluyentes o no se pueden realizar (por ejemplo, en ausencia de bucles cerrados en la red), la evaluación del supuesto de transitividad se basa en la consideración no estadística de la similitud de las características modificadoras de efecto, y esta información se debe describir con suficiente detalle.

¿Cómo evaluar esta afirmación?

Es posible que los meta-analistas hayan abordado el problema con métodos estadísticos apropiados (p. ej., metarregresión, análisis de subgrupos o modelos de efectos mixtos) que probablemente explicaran estas diferencias, o que el análisis de sensibilidad haya demostrado que las diferencias eran irrelevantes. Los modificadores potenciales del efecto podrían haberse medido a nivel del participante individual o a nivel del estudio. Idealmente, las características del participante que modifican el efecto potencial se habrían medido a nivel individual. Una metarregresión de red adecuada requiere entonces datos de participantes individuales. Las características del participante podrían haberse agregado a nivel de estudio, pero este hallazgo introduce un riesgo de sesgo de agregación.

V: si los autores han utilizado métodos aceptables para realizar el análisis, como métodos estadísticos (por ejemplo, metarregresión, análisis de subgrupos o modelos de efectos mixtos) que probablemente explicaran estas diferencias, o el análisis de sensibilidad mostró que las diferencias no eran importantes en la revisión, ajustándose al conjunto inclusivo de modificadores de efectos en cuestión, y abordó la comparabilidad de los hallazgos con los de los análisis no ajustados.

PV: si los análisis se exploraron parcialmente, pero persisten diferencias que no se abordaron o si los hallazgos de los análisis no se consideraron adecuadamente.

F o PF: si no se exploraron análisis para observar las diferencias en potenciales modificadores del efecto.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

Los evaluadores deben decidir qué riesgo atribuir a la revisión (alto/bajo/algunas preocupaciones) con la cita textual del manuscrito y su justificación para su elección.

Dominio 3: síntesis estadística

El último dominio se centra en la síntesis estadística, abarcando aspectos como los sesgos de no reporte, los sesgos dentro de los estudios, los métodos estadísticos, la inconsistencia entre la evidencia directa e indirecta (incoherencia), y los análisis de sensibilidad. La afirmación 3.8 se considera sólo si la afirmación 3.7 se considera F o PF. En este dominio, los autores de MAR pueden utilizar el análisis de sensibilidad para abordar las deficiencias del análisis principal que darían lugar a una calificación de PV, PF o F. Si se realizó un análisis de sensibilidad que habría recibido una calificación más alta, se puede asignar dicha calificación si el análisis de sensibilidad arrojó resultados similares a los resultados principales.

3.1 El análisis respetado dentro de la aleatorización del estudio

Un método estadístico simple pero inapropiado es comparar los resultados de brazos individuales de diferentes estudios como si fueran de un solo estudio. Un análisis que ignora la aleatorización dentro del estudio utiliza implícitamente comparaciones entre estudios, así como dentro de los estudios. Una comparación indirecta no

ajustada, que combina datos de ensayos como si los datos provenieran de un ensayo grande, ignora la aleatorización dentro del estudio y utilizaron un enfoque de comparación indirecta no ajustada. El análisis de ECAs siempre debe basarse en comparaciones entre grupos comparables. Por lo tanto, un MAR de ECAs también debe extraer conclusiones basadas en la evidencia aleatorizada (es decir, en las comparaciones intra-estudio). Un MAR de estudios observacionales también debe basarse en comparaciones intra-estudio, ya que es probable que estas hayan intentado controlar los factores de confusión.

Fuente de sesgo

Las comparaciones entre estudios pueden estar sesgadas incluso por pequeñas diferencias en los resultados generales entre estudios, que surgen, por ejemplo, de pequeñas diferencias en los grupos de pacientes o en las definiciones de resultados.

¿Cómo evaluar esta afirmación?

Combinar grupos de estudios (para cada intervención) y comparar los resultados generales de los grupos en las intervenciones de la red no respeta la aleatorización y sería un enfoque inadecuado. Los estudios más antiguos pueden ser más propensos a este enfoque porque los métodos para los MAR no estaban completamente desarrollados ni popularizados. Cualquier análisis que utilice comparaciones dentro del estudio como datos respeta la aleatorización. El análisis de los datos a nivel de grupo o de paciente puede o no preservar la aleatorización. El análisis respeta la aleatorización si se incluye un efecto fijo del estudio en el modelo estadístico. Los métodos de análisis basados en grupos incluyen un efecto aleatorio del estudio en el modelo estadístico y pueden utilizar comparaciones entre estudios. Es probable que este enfoque genere sesgo si los resultados promedio difieren en las comparaciones directas de la red.

V: si se analizaron comparaciones dentro del estudio o se analizaron datos a nivel de brazo o de paciente con un modelo que incluye un efecto fijo del estudio.

PV: si los datos a nivel de brazo o de paciente se analizaron con un modelo que incluye un efecto aleatorio del estudio y los resultados promedio fueron similares en las comparaciones directas. También es probable que sea verdadera si este enfoque se utilizó en un análisis de sensibilidad que arrojó resultados similares a los resultados principales.

PF: si los datos a nivel de brazo o de paciente se analizaron con un modelo que incluye un efecto aleatorio del estudio y los resultados promedio fueron diferentes en las comparaciones directas.

F: si se analizaron datos a nivel de brazo o de paciente ignorando un modelo que incluye un efecto de estudio aleatorio o fijo, o si los datos se combinaron entre estudios antes de compararlos entre intervenciones.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

3.2 No se sospechó sesgo de publicación ni otros sesgos selectivos de no reporte

Como cualquier meta-análisis, un MAR debe incluir todos los estudios que han recopilado datos relevantes para la pregunta planteada. La falta de estudios completos, resultados específicos (p. ej., desenlaces) y análisis desfavorables de un MAR pueden verse influenciados por el valor de p, la dirección o la magnitud de los efectos del estudio. Un ejemplo típico es que los resultados significativos que sugieren la eficacia de un tratamiento tienen mayor probabilidad de publicarse que los resultados no significativos. Los estudios con resultados significativos también tienen mayor probabilidad de estar disponibles rápidamente, publicarse en revistas de alto impacto y ser citados por otros y, por lo tanto, ser más fácilmente identificables para revisiones sistemáticas. El sesgo en la selección del resultado informado, donde los autores del estudio seleccionan un resultado para informar entre múltiples mediciones o análisis, basándose en el valor de p, la magnitud o la dirección de los resultados, es diferente del sesgo debido a la falta de resultados. El término sesgo de reporte se ha utilizado a menudo para describir este problema, pero aquí utilizaremos el término sesgo de no reporte.

Un estudio, resultado o análisis puede no informarse por varias razones:

- se realizó un estudio, pero no se publicó;
- el resultado relevante de un estudio incluido no está disponible para los autores de la revisión;
- los autores de la revisión no han recopilado ni procesado, involuntariamente, los datos disponibles;

- los autores de la revisión han excluido intencionalmente el resultado o un análisis del MAR.

Fuente de sesgo

La falta de estudios o resultados primarios puede causar sesgo en las estimaciones del efecto de la intervención. La consideración importante aquí es si el resultado del estudio no se informó debido al hallazgo (p. ej., magnitud, dirección o valor P del resultado), en cuyo caso la omisión del resultado del estudio introducirá sesgo en el MAR. Este sesgo a menudo se conoce como sesgo de publicación cuando los estudios no se publican, o (dentro de los estudios) sesgo de informe selectivo cuando se ocultan resultados o análisis no deseados y no se publican. Si un resultado se excluye por razones no relacionadas con el hallazgo (p. ej., únicamente por razones prácticas, como la extensión del artículo), el MAR puede perder precisión, pero no debería correr el riesgo de sesgo. Los autores de la revisión pueden tomar decisiones inapropiadas para excluir algunos resultados del MAR. Un ejemplo potencial sería la exclusión de estudios considerados responsables de un gran grado de heterogeneidad entre estudios basándose únicamente en consideraciones estadísticas (aunque estas exclusiones pueden ser razonables en un análisis de sensibilidad).

¿Cómo evaluar esta afirmación?

Determinar si la revisión sistemática no identificó estudios es un desafío. Los gráficos de embudo ajustados por comparación y los análisis estadísticos relacionados pueden brindar indicios sobre la posibilidad de sesgo de notificación dentro de los estudios y entre ellos. El evaluador también debe examinar el número de estudios identificados para su posible inclusión en el MAR (p. ej., a partir de un diagrama de flujo dentro de una revisión sistemática subyacente) y el número de estudios realmente incluidos en el MAR, junto con las razones para la exclusión de estudios o resultados. Una discrepancia en los números daría lugar a que se considerara por qué no fueron incluidos resultados relevantes en el MAR.

V: si era probable que se hubieran identificado todos los estudios y todos los resultados relevantes de los estudios identificados se incluyeron en el MAR o está claro que cualquier resultado no incluido en el MAR fue excluido por razones no relacionadas con los hallazgos (por ejemplo, porque los estudios no midieron el resultado de interés).

PV: si la cantidad de evidencia potencialmente excluida del MAR es tan pequeña que su inclusión tendría un efecto trivial en los hallazgos, o si algún resultado no incluido en el MAR probablemente fue excluido por razones no relacionadas con los hallazgos.

PF: si es probable que falten resultados elegibles en el MAR (debido a la supresión por parte de los autores del estudio o la exclusión por parte de los autores de la revisión) y pueden ser sistemáticamente diferentes de los resultados incluidos en el MAR o la cantidad de evidencia que falta en el MAR es suficientemente grande como para que su inclusión pueda afectar los efectos de la intervención estimados del MAR.

F: si existe evidencia de que faltaban resultados elegibles en el MAR (debido a la supresión por parte de los autores del estudio o la exclusión por parte de los autores de la revisión) y es probable que estos resultados sean sistemáticamente diferentes de los resultados incluidos en el MAR, con una cantidad suficiente de evidencia faltante que podría afectar los efectos de intervención estimados del MAR.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

3.3 Sólo se informaron todos los análisis predefinidos, o se explicaron las discrepancias

El MAR debería haber seguido un protocolo publicado o accesible o un plan de análisis estadístico. Algunos ejemplos son los protocolos registrados en la base [PROSPERO](#), los protocolos publicados en línea o los protocolos depositados en un repositorio en línea. Un protocolo podría evitar la toma de decisiones subjetivas sobre cómo se realiza el MAR y qué datos se incluyen. Sin un protocolo, es probable que las decisiones sobre qué estudios se incluyen, qué datos se extraen y cómo se analizan los datos se tomen a la luz de los datos. Con los diversos resultados informados para cada ensayo incluido en un meta-análisis, los autores podrían seleccionar deliberadamente un resultado favorable o desfavorable de cada estudio. Al seleccionar los mejores resultados de esta manera, los revisores que eligen los estudios que se incluirán podrían sesgar los hallazgos hacia una suposición preconcebida. Por ejemplo, en un estudio de métodos que examina los resultados de los efectos del

fármaco gabapentina para el dolor, los autores podrían manipular los hallazgos de tal manera que los meta-análisis con los efectos más extremos en cualquier dirección hicieran que el fármaco pareciera altamente efectivo o completamente ineficaz para tratar el dolor neuropático. Aunque los métodos descritos en un protocolo generalmente deben seguirse, a veces es necesario desviarse de los métodos planificados. Cualquier cambio de este tipo debe realizarse de manera que no genere sesgos y debe explicarse detalladamente.

Fuente de sesgo

Los autores de la revisión pueden introducir sesgos a través de su selección de análisis, métodos de análisis y datos, de modo que se oculten o reemplacen resultados desfavorables.

¿Cómo evaluar esta afirmación?

Se debe obtener el protocolo (es decir, el informe planificado previamente o registrado) y leerlo junto con la MAR para determinar si se han omitido análisis.

V: si se siguieron claramente los análisis predefinidos (por ejemplo, porque se siguió un protocolo detallado y pre establecido).

PV: si hay una indicación de que se siguieron análisis predefinidos, y solo esos análisis (por ejemplo, porque un protocolo no está lo suficientemente detallado pero los métodos implementados son sensatos y de acuerdo con lo que estaba escrito en el protocolo), no hay ningún protocolo disponible pero la sección de métodos parece rigurosa y todos los análisis mencionados se abordan en los resultados, o todas las desviaciones de los métodos planificados se justificaron con razones no relacionadas con los resultados observados.

PF: si no hay ninguna indicación de que se hayan seguido análisis predefinidos, por ejemplo, debido a detalles insuficientes sobre los métodos que probablemente se hayan planificado e implementado (por ejemplo, falta de protocolo y claridad insuficiente en la sección de métodos).

F: si claramente no se siguieron los análisis predefinidos, o se utilizaron otros análisis, y hay una indicación (o evidencia) de que las desviaciones de los análisis predefinidos se realizaron debido a los resultados.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

3.4 Los sesgos en los estudios primarios fueron mínimos o se abordaron en la síntesis

Los MAR que incluyen estudios con alto riesgo de sesgo también presentan riesgo de sesgo, lo que puede llevar a una sobreestimación o subestimación de las estimaciones del efecto de la intervención. Por lo tanto, es importante evaluar el potencial de sesgo en los resultados de los estudios primarios incluidos. En un MAR, es fundamental comprender cómo los estudios que informan contrastes específicos afectan el sesgo en los efectos de la intervención para otros contrastes. Comprender cómo los estudios que informan contrastes específicos dependen de la estructura de la red y puede ser complejo de evaluar por completo.

Fuente de sesgo

Las fallas en la realización de ensayos aleatorios y otros tipos de estudios primarios pueden dar lugar a una estimación sesgada del efecto de la intervención en una red de intervenciones.

¿Cómo evaluar esta afirmación?

Los autores de la revisión deberían haber evaluado el riesgo de sesgo con una herramienta apropiada, como la herramienta Cochrane RoB 2.0 para ECAs. En ocasiones, se pueden extraer conclusiones válidas de los MAR que incluyen resultados que se consideran de alto riesgo de sesgo, lo que se debe en parte a que un estudio de alto riesgo de sesgo puede no arrojar resultados sesgados y en parte a que en ocasiones se puede establecer que estudios concretos no tienen efecto sobre las conclusiones. Los análisis de sensibilidad o los análisis de umbral pueden utilizarse para demostrar, por ejemplo, que la exclusión de estudios de alto riesgo de sesgo no modifica los efectos estimados de la intervención. Como alternativa, existen métodos para ajustar el sesgo en los MAR. Los principales enfoques son los que incorporan ajustes directos para el sesgo en estudios individuales (p. ej., con distribuciones previas para sesgos) o son específicos del contexto de un MAR y pueden estimar y ajustar el

sesgo solo con los datos incluidos. Si los autores del MAR evaluaron la certeza en su conjunto de evidencia (por ejemplo, con GRADE-MAR o CINeMA), esta evaluación no justifica una respuesta V o PV.

V: si todos los estudios fueron evaluados como de bajo riesgo de sesgo o los análisis de sensibilidad mostraron que incluir estudios con mayor riesgo de sesgo no tuvo efecto sobre los resultados.

PV: si la proporción de información con alto riesgo de sesgo fue demasiado pequeña como para afectar los resultados, los análisis de sensibilidad o de umbral mostraron que incluir estudios con alto riesgo de sesgo tuvo un efecto mínimo, o se utilizaron enfoques de ajuste que probablemente hayan corregido los sesgos.

PF: si la proporción de información con alto riesgo de sesgo fue suficiente para afectar los resultados y los análisis de sensibilidad no mostraron que la inclusión de estudios con alto riesgo de sesgo tuviera un efecto mínimo, se utilizaron enfoques de ajuste de sesgo y no está claro si es probable que hayan corregido los sesgos, o no se evaluó el riesgo de sesgo en los estudios incluidos.

F: si existen sesgos importantes en los estudios primarios y estos no han sido abordados por los revisores o los análisis de sensibilidad muestran que los resultados están fuertemente influenciados por estudios con mayor riesgo de sesgo.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

3.5 Se utilizaron métodos apropiados para manejar estudios de múltiples brazos

Si un MAR incluye estudios que comparan (individualmente) tres o más intervenciones (estudios multibrazo o de múltiples ramas), es importante garantizar que se gestionen adecuadamente. Dos aspectos pueden introducir sesgo: cómo se representan los estudios multibrazo en los datos y cómo se analizan en el modelo estadístico. La principal preocupación en este último caso es cómo se modelan los efectos aleatorios.

Al manejar datos a nivel de estudio, el MAR puede incorporar evidencia de estudios multibrazo al conservar la estructura multibrazo original del estudio y considerar adecuadamente la correlación de los efectos entre los brazos. Cuando se utilizan datos a nivel de brazo en el análisis (normalmente solo se realiza en el marco bayesiano), es improbable que la inclusión de múltiples brazos por separado sea una fuente de sesgo. Una alternativa es generar comparaciones por pares dentro de los estudios multibrazo antes de realizar el MAR e incluir estas comparaciones por pares (contrastos) en el conjunto de datos. Incluir múltiples contrastes de un estudio en un modelo basado en contrastes no será una fuente de sesgo si se consideran explícitamente las correlaciones entre dichos contrastes. Sin embargo, si se ignoran estas correlaciones, las intervenciones que ocurren en más de un contraste tendrán una influencia injustificada en el MAR. Cuando se utiliza un modelo de efectos aleatorios, el método de análisis debe considerar adecuadamente la correlación entre los múltiples efectos aleatorios que se estiman a partir de un estudio multibrazo.

Fuente de sesgo

Los brazos de un estudio multibrazo están correlacionados y no son independientes. Si los autores del análisis multibrazo ignoraran la dependencia de los brazos, las varianzas de las estimaciones del efecto se subestimarían y los sesgos en las estimaciones del efecto podrían exagerarse debido a ponderaciones incorrectas. Ignorar la correlación probablemente también subestimaría la inconsistencia de toda la red. Sin embargo, una consideración incorrecta de las correlaciones podría resultar en un cálculo incorrecto de los intervalos de confianza o de credibilidad, lo que podría generar sesgos en las conclusiones.

¿Cómo evaluar esta afirmación?

Se debe evaluar tanto el manejo de los datos a nivel de estudio como el modelado de efectos aleatorios. Para el manejo de los datos a nivel de estudio, la información sobre cómo se extrajeron e incluyeron en la red los datos del estudio primario multibrazo debe estar disponible en la sección de análisis de datos del informe del MAR. Si un estudio multibrazo se convierte en un estudio de dos brazos mediante la fusión de los brazos activos en uno solo, es improbable que esta acción genere sesgo, siempre que el brazo fusionado corresponda a un nodo de intervención en la red (véase el punto 1.3). Si un brazo de intervención de control se divide para formar varios estudios de dos brazos, es probable que esta acción introduzca sesgo en los errores estándar de los resultados del MAR, ya que un

cálculo incorrecto de los errores estándar podría tener implicaciones en las ponderaciones y, por lo tanto, en la estimación puntual o los intervalos de confianza obtenidos. Cuando los datos están en formato de nivel de contraste, el evaluador debe verificar si los contrastes se formaron sin dividir los brazos de control y si los meta-analistas especificaron correctamente las correlaciones en los contrastes múltiples de los ensayos multibrazo utilizados en el modelo de síntesis. Este hallazgo podría ser especialmente el caso si los ensayos de múltiples brazos aportan la mayor parte de la evidencia en una red o en comparaciones particulares, como puede ser el caso si muchos de los estudios están explorando diferentes dosis.

Al modelar efectos aleatorios, es importante que el método de análisis elegido considere la correlación de los efectos aleatorios en estudios multibrazo. Esta correlación se aplica tanto a modelos basados en brazos como a modelos basados en contraste donde se utilizan modelos de efectos aleatorios. El software o código estadístico puede incorporar estos métodos para gestionar estudios multibrazo, como el código WINBUGS reportado por Dias et al. Varios paquetes de R ofrecen funcionalidad para los MAR de ensayos (MAR), incluyendo la gestión de ensayos multibrazo. Los paquetes más populares incluyen gemtc, netmeta, netmetaXL y rjags (para su uso con JAGS). (Otro ejemplo de Gibbs). Esta lista no es exhaustiva; quienes no estén familiarizados con paquetes similares deberían consultar con un estadístico.

V: si no existen estudios multibrazo, como se indica claramente en las descripciones de los estudios incluidos. También si los datos de los estudios multibrazo se incluyeron como datos a nivel de grupo o como efectos relativos, con una correlación adecuada, y los métodos utilizados consideran adecuadamente la correlación en los efectos aleatorios de los estudios multibrazo (si se utilizó un modelo de efectos aleatorios).

PV: si no se informó completamente el software, el código o los métodos de MAR para abordar la correlación de los efectos aleatorios en estudios de múltiples brazos, pero otra evidencia sugiere que se utilizaron el software y los métodos apropiados.

F o PF: si los estudios de múltiples brazos no estuvieron representados adecuadamente en el conjunto de datos o los métodos elegidos no tuvieron en cuenta adecuadamente las correlaciones en los efectos aleatorios o no indicaron el manejo apropiado de estas correlaciones.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

3.6 Se hicieron suposiciones apropiadas sobre la homogeneidad o heterogeneidad de los efectos dentro de las comparaciones

En un MAR se pueden hacer diversas suposiciones sobre la heterogeneidad de los efectos de una intervención. Por heterogeneidad nos referimos aquí a la variación en los efectos de una intervención dentro de una comparación específica. Se pueden considerar cuatro enfoques:

- asumir que el efecto de la intervención subyacente es el mismo (es homogéneo) para todos los estudios de una comparación particular, y que este efecto subyacente se aplica a cada comparación (modelo MAR de efecto común o de efecto fijo);
- asumir que existe heterogeneidad de efectos dentro de las comparaciones y forzar que la cantidad de heterogeneidad sea la misma para cada comparación (modelo de efectos aleatorios restringidos);
- Supongamos que puede existir heterogeneidad de efectos dentro de las comparaciones y permitamos esta heterogeneidad de efectos sea diferente para diferentes comparaciones (modelo de efectos aleatorios flexible); y
- suponer que el efecto de la intervención subyacente es el mismo para todos los estudios de algunas comparaciones, pero permitir la heterogeneidad del efecto entre los estudios de otras comparaciones (modelo híbrido).

Los enfoques más comunes son el modelo de efectos fijos y el modelo de efectos aleatorios restringidos. Este último requiere la estimación de una sola cantidad para representar la heterogeneidad. Si la suposición hecha sobre la heterogeneidad es incorrecta, los resultados del MAR pueden verse afectados. Si se asume erróneamente que la heterogeneidad está ausente, no se reflejará en las inferencias, que pueden parecer más precisas de lo que deberían ser. Si se permite la heterogeneidad, pero se asume erróneamente que es la misma para todas las comparaciones, las inferencias para las comparaciones con mayor heterogeneidad (verdadera) pueden parecer más

precisas de lo que serían con un modelo de efectos aleatorios flexible, mientras que las inferencias para las comparaciones con menor heterogeneidad (verdadera) pueden parecer menos precisas de lo que serían. Con pocos estudios para cada comparación, la heterogeneidad puede ser mal estimada, lo que a su vez podría inflar la incertidumbre.

Las discrepancias entre los supuestos y los datos afectan la ponderación asignada a los estudios en el MAR y, por lo tanto, a algunas fuentes de evidencia indirecta se les pueden asignar ponderaciones inapropiadamente altas o bajas. Si existe heterogeneidad inexplicable, se puede utilizar la metarregresión para investigar si esta se explica por las características del nivel de estudio o de la intervención. Por ejemplo, la heterogeneidad entre estudios que implementan un fármaco en diferentes dosis se puede explicar incluyendo la dosis como covariable en una metarregresión de red. Si la heterogeneidad se explica adecuadamente mediante la metarregresión, la elección entre diferentes modelos de efectos fijos y aleatorios será menos importante.

Fuente de sesgo

La heterogeneidad en un MAR puede resultar en diferentes estimaciones de los efectos de la intervención cuando se eligen diferentes supuestos de homogeneidad o heterogeneidad. Los modelos que otorgan diferentes esquemas de ponderación a estudios grandes y pequeños en el MAR pueden exagerar los sesgos (p. ej., los de estudios más pequeños). Por lo tanto, los supuestos inapropiados de homogeneidad o heterogeneidad implican un riesgo de sesgo, ya que los estudios pueden haber sido seleccionados por los autores del MAR con base en los resultados. Los supuestos inapropiados de homogeneidad o heterogeneidad a menudo también resultan en estimaciones incorrectas de los intervalos de confianza o de credibilidad. Los supuestos inapropiados de homogeneidad o heterogeneidad deben tenerse en cuenta al evaluar el sesgo en las conclusiones del MAR.

¿Cómo evaluar esta afirmación?

La información sobre las suposiciones realizadas sobre la heterogeneidad en el modelo de análisis debe estar disponible en la sección de análisis de datos del informe de la MAR. También es útil examinar la naturaleza de la heterogeneidad entre los estudios identificados, la cual puede obtenerse mediante diagramas de bosque (*forest plot*) de comparaciones directas. Con frecuencia, se presentarán comparaciones directas con un gran número de estudios (que brindan información sobre la heterogeneidad) y otras comparaciones con pocos estudios (que generalmente no la brindan). Se debe hacer hincapié en las comparaciones con un mayor número de estudios.

V o PV: si no se modela la heterogeneidad (p. ej., con un modelo de efectos fijos para eventos poco frecuentes) y este enfoque está bien justificado, o si se modela la heterogeneidad inexplicable (es decir, con un modelo de efectos aleatorios) y no existe evidencia en los datos que contradiga los supuestos del modelo. Una justificación para un enfoque de modelado particular es si se realiza un análisis de sensibilidad con un enfoque alternativo igualmente apropiado o más general, y este análisis de sensibilidad muestra resultados similares a los resultados principales.

F o PF: si no se modela la heterogeneidad y este enfoque no está bien justificado, o si se modela la heterogeneidad y existe evidencia en los datos en contra de los supuestos del modelo.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

3.7 No hay evidencia de conflicto entre las estimaciones directas e indirectas del mismo efecto

Un MAR no es válido cuando la evidencia directa e indirecta no concuerda (incoherencia). Un conflicto entre la evidencia directa e indirecta implica que el supuesto fundamental de transitividad en el MAR no se cumple. Explorar la evidencia de dicho conflicto es una parte importante de un MAR. Las afirmaciones 2.1-2.3 analizaron si la naturaleza de los estudios hace plausible la transitividad. Esta afirmación responde si los efectos de la intervención ofrecen alguna evidencia sobre la transitividad.

La inconsistencia o mejor llamada incoherencia (la manifestación estadística de la transitividad) solo puede evaluarse cuando existen bucles cerrados en la red. Un bucle cerrado se define como un conjunto de contrastes para el cual se dispone de fuentes independientes de evidencia directa e indirecta. Los bucles formados únicamente por comparaciones de un estudio multibrazo son inherentemente consistentes y no requieren examen, ya que la transitividad debe mantenerse dentro de un estudio multibrazo.

La evidencia sobre conflictos entre fuentes directas e indirectas de evidencia puede provenir de investigaciones locales de inconsistencia (p. ej., división de nodos) o investigaciones globales de inconsistencia (p. ej., comparaciones de ajuste de modelos de consistencia e inconsistencia a los datos, prueba de diseño por tratamiento). Una forma problemática de identificar conflictos es comparar estimaciones indirectas con estimaciones de MAR, porque la evidencia indirecta está incluida en ambas estimaciones. Comparar estimaciones indirectas con estimaciones directas es mejor porque las estimaciones son estadísticamente independientes. En algunas redes, pueden surgir sospechas de que un bucle cerrado particular o una región particular de la red tiene una mayor probabilidad de inconsistencia. En este caso, la evaluación del conflicto debe incluir una o más evaluaciones locales para ese bucle o región. De lo contrario, la inconsistencia se puede evaluar en toda la red con múltiples métodos locales o un método global.

Fuente de sesgo

Cuando la evidencia directa e indirecta para una comparación no concuerda, dicha comparación contiene información inconsistente. Dependiendo de qué fuente de evidencia se acerque más a la verdad para esta comparación en relación con la pregunta de investigación subyacente planteada por la red, y de la cantidad relativa de información en cada fuente, los resultados de esta comparación y de otras comparaciones basadas en ella pueden estar sesgados.

¿Cómo evaluar esta afirmación?

Para determinar si existe evidencia de inconsistencia, es importante considerar los métodos utilizados para identificarla. Los evaluadores deben examinar si los meta-analistas analizaron adecuadamente los bucles cerrados y si los resultados se interpretaron considerando la estructura de la red, la dirección de los efectos y la incertidumbre de las estimaciones. Se ha demostrado que el método global tiene poca potencia para detectar inconsistencias.

V: si no existen fuentes potenciales de inconsistencia en la red (es decir, no hay bucles cerrados o bucles formados solo por las comparaciones en un estudio de múltiples brazos), o si se aplicó una exploración adecuada de la inconsistencia sin ninguna indicación de que hubiera inconsistencia presente (este enfoque generalmente requeriría una base de evidencia considerable para superar los problemas de bajo poder en las pruebas de inconsistencia).

PV: si se realizó una exploración adecuada de la inconsistencia y no existe evidencia de ella. Este enfoque podría incluir casos en los que algunas estimaciones son extremas debido a la presencia de celdas cero, por ejemplo, pero la dirección general del efecto es consistente.

PF: si existe alguna evidencia de inconsistencia, pero no fue grave, o si surgieron preocupaciones importantes de que las pruebas realizadas tenían bajo poder para detectar inconsistencia.

F: si no se proporciona información sobre la inconsistencia o existe evidencia clara de inconsistencia.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

3.8 Si es falso o probablemente falso el punto 3.7: los resultados contradictorios entre la evidencia directa e indirecta se abordaron adecuadamente

La incoherencia entre la evidencia directa e indirecta puede generar sesgo en los resultados de la red. Una forma adecuada de evitar el sesgo puede incluir una revisión cuidadosa de las características del estudio y la población (incluido el riesgo de sesgo y la definición de nodos) para identificar diferencias en los estudios que comparan diferentes intervenciones. Se pueden considerar análisis alternativos, como el análisis de subgrupos o el análisis de sensibilidad, la agrupación o división de nodos, o la consideración de supuestos alternativos (p. ej., ajustar un modelo de efectos aleatorios en lugar de un modelo de efectos fijos o ajustar las covariables mediante metarregresión). Sin embargo, estos análisis deben considerarse exploratorios. Además, estos pasos deben tener una justificación adecuada y los análisis adicionales deben estar descritos en el protocolo.

¿Cómo evaluar esta afirmación?

El evaluador puede observar cómo los meta-analistas examinaron los conflictos entre la evidencia directa e indirecta y juzgar si se exploraron adecuadamente. Un enfoque particularmente problemático consiste en omitir estudios

primarios individuales hasta que se elimine la inconsistencia (es decir, hasta que el valor p de la inconsistencia caiga por debajo de un umbral de significación estadística). Los métodos para abordar la inconsistencia incluyen el uso de un modelo que la permita explícitamente (como el modelo de interacción de diseño por tratamiento y el modelo de metarregresión en red). El enfoque adecuado dependerá de la naturaleza y el alcance de la inconsistencia, de los datos disponibles y de la pregunta de investigación planteada.

V o PV: si los análisis alternativos tenían una justificación clara y resolvieron el conflicto original.

F o PF: si los resultados del meta-análisis directo se compararon con los resultados de MAR o literatura previa (por ejemplo, revisiones sistemáticas) para abordar resultados conflictivos, o si los estudios primarios se excluyeron aleatoriamente hasta que se alcanzó la consistencia.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

3.9 Si se realizó un análisis bayesiano, la elección de distribuciones previas fue apropiada

Este ítem no es aplicable si no se realizó un análisis bayesiano (es decir, análisis frecuentista).

El análisis bayesiano requiere distribuciones previas para parámetros desconocidos, incluyendo la varianza o varianzas de heterogeneidad y los efectos de la intervención. Estas distribuciones previas pueden ser informativas o vagas (no informativas). Las distribuciones previas informativas pueden basarse en conocimiento previo (p. ej., con distribuciones empíricas para especificar una distribución previa para la varianza de heterogeneidad o con datos observacionales externos al MAR para especificar una distribución previa para los efectos de la intervención). La elección de distribuciones previas debe justificarse y ser apropiada, particularmente cuando solo existen unos pocos estudios para cada comparación o pocas conexiones (comparaciones) y bucles en la red. Si se utiliza una distribución previa informativa, se debe indicar y justificar la fuente de la información. Una distribución previa informativa sobre los efectos de la intervención sólo puede justificarse en circunstancias excepcionales.

Fuente de sesgo

Las distribuciones previas inapropiadas pueden causar sesgo, especialmente cuando la evidencia es escasa, si representan suposiciones incorrectas (p. ej., si son demasiado restrictivas) o si incorporan evidencia o creencias falsas o poco realistas. Algunas distribuciones previas inapropiadas (p. ej., para la varianza de la heterogeneidad) pueden generar intervalos de confianza o de credibilidad incorrectos, aunque no estimaciones puntuales, lo que puede sesgar las conclusiones.

¿Cómo evaluar esta afirmación?

Cuando se afirma que las distribuciones previas no son informativas, el evaluador puede considerar evaluar el rango de distribuciones previas en relación con las expectativas previas para los parámetros y la cantidad de información en la red, según lo informado por los meta-analistas. El evaluador puede examinar las densidades posteriores, si se informan, para garantizar que no haya truncamiento artificial y que estén suficientemente informadas por los datos. Cuando las distribuciones previas son informativas, el evaluador puede considerar si se basan en datos relevantes para el MAR en consideración. Por ejemplo, las distribuciones previas informativas para los parámetros de heterogeneidad deben basarse en evidencia de estudios similares en términos de comparaciones de intervenciones y resultados.

V: si se utilizaron distribuciones previas que claramente no eran informativas en todo el rango de valores de parámetros posibles, si se utilizaron distribuciones previas informativas y tenían una justificación clara, o si se realizó un análisis de sensibilidad con distribuciones previas alternativas apropiadas y mostró resultados similares a los resultados principales.

PV: si se utilizaron distribuciones previas no informativas disponibles comercialmente.

F: o probablemente falso si se utilizaron distribuciones previas informativas que no estaban claramente justificadas.

SI: cuando no se proporcionan suficientes datos para permitir que se emita un juicio razonable.

Los evaluadores deben decidir qué riesgo atribuir a la revisión (alto/bajo/algunas preocupaciones) con la cita textual del manuscrito y su justificación para su elección.

Riesgo general de sesgo en el MAR

Un juicio global puede hacerse a nivel de los resultados o conclusiones del MAR, o ambos. Cuál de estos le interesa al evaluador dependerá del propósito de la evaluación del riesgo de sesgo y de si el evaluador utilizó los resultados o las conclusiones de la revisión. Para el juicio global, el riesgo de sesgo a nivel de revisión sistemática (p. ej., con una herramienta apropiada, como ROBIS (o AMSTAR 2) se combina con los juicios de la herramienta RoB NMA a nivel de dominio para determinar si los hallazgos del MAR en su conjunto están en riesgo de sesgo. Al fusionar una evaluación ROBIS con una evaluación RoB NMA, el evaluador tendrá que considerar sus juicios de los primeros tres dominios de la herramienta ROBIS, así como los tres juicios a nivel de dominio de la herramienta RoB NMA. La **Figura 1** presenta un formato tabular sugerido para una evaluación global del riesgo de sesgo y el apéndice C tiene un segundo formato sugerido. La fase 3 de la herramienta ROBIS se omite cuando la herramienta RoB NMA se utiliza con la herramienta ROBIS. La herramienta ROBIS se realiza a nivel de revisión sistemática, mientras que la herramienta RoB NMA se utiliza para cada MAR por separado, por lo que, en una revisión sistemática con MAR, se realizará una evaluación ROBIS, pero múltiples evaluaciones RoB NMA.

Figura 1. Riesgo general de sesgo MAR integrando ROBIS y RoB NMA

| MAR dentro de una RS | Dominios de ROBIS | | | Dominios de RoB NMA | | | Juicios globales | |
|----------------------|---|--|---|--------------------------------------|-------------------------|----------------------|------------------|--------------|
| | Criterios de elegibilidad de los estudios | Identificación y selección de los estudios | Recolección de los datos y evaluación de los estudios | Intervenciones y geometría de la red | Modificadores de efecto | Síntesis estadística | Resultados | Conclusiones |
| ID 1 | + | ? | - | - | ? | - | - | - |
| ID 2 | ? | ? | - | + | ? | - | - | - |
| ID 3 | + | + | + | ? | - | + | - | + |
| ID 4 | + | + | ? | + | + | + | + | + |
| ID 5 | + | + | ? | ? | + | + | + | - |

MAR: Metanálisis en red. RS: revisión sistemática: ROBIS: herramienta para evaluar el riesgo de sesgo en revisiones sistemáticas. RoB NMA: Riesgo de sesgo en meta-análisis en red.

Con base en las respuestas a las declaraciones de señalización en cada uno de los dominios, un evaluador primero toma una decisión a nivel de dominio sobre el riesgo de sesgo en los resultados del MAR en general. Si todos los dominios se juzgan como de bajo riesgo, el resultado del MAR en su conjunto puede juzgarse como de bajo riesgo. Si se asignó un alto riesgo de sesgo a cada uno de los cinco dominios, el resultado del MAR generalmente debe juzgarse como de alto riesgo de sesgo. Algunas preocupaciones en múltiples dominios pueden hacer que los evaluadores decidan sobre un juicio general de alto riesgo de sesgo para ese resultado. El riesgo de sesgo en los resultados debe ser el enfoque si los resultados del MAR se utilizan en modelos de decisión, pero el riesgo de sesgo en las conclusiones debe ser el enfoque si las conclusiones del MAR se van a utilizar en la toma de decisiones.

Sesgo en los resultados de MAR

En esta sección, se solicita a los evaluadores que consideren si existe un posible sesgo en los efectos estimados de la intervención (es decir, los resultados del MAR). Las respuestas para esta evaluación son bajo riesgo de sesgo, alto riesgo de sesgo o algunas preocupaciones. Si todos los dominios se evaluaron con bajo riesgo de sesgo, generalmente se debe emitir una evaluación de bajo riesgo de sesgo. De lo contrario, el evaluador debe decidir entre algunas preocupaciones y alto riesgo de sesgo, según la combinación de respuestas a las afirmaciones de señalización.

Sesgo en las conclusiones de la MAR

Para evaluar el sesgo en las conclusiones de una revisión sistemática con un MAR, se solicita a los evaluadores que asignen un juicio de si existen o no preocupaciones. Por lo tanto, el evaluador debe determinar si, en sus conclusiones, los autores del MAR abordaron todas las limitaciones identificadas en la evaluación. Los puntos 3.5, 3.6 y 3.8 deben reconsiderarse al evaluar el riesgo general de sesgo en las conclusiones, ya que las elecciones de modelado inadecuadas identificadas en estos puntos pueden generar incertidumbre en los resultados, subestimados o sobreestimados, lo que probablemente conducirá a conclusiones sesgadas en el MAR. Al evaluar el sesgo en las conclusiones, la importancia que se le da a las diferentes limitaciones depende del contexto del MAR que se evalúa.

Los efectos estimados de la intervención podrían presentar un alto riesgo de sesgo, mientras que las conclusiones del MAR podrían presentar un bajo riesgo de sesgo. Si los efectos estimados de la intervención presentan un alto riesgo de sesgo debido a la naturaleza de los estudios primarios, pero este riesgo se tiene cuidadosamente en cuenta al extraer conclusiones, las conclusiones del MAR podrían presentar un bajo riesgo de sesgo. Sin embargo, si los efectos estimados de la intervención presentan un alto riesgo de sesgo debido a la realización del MAR, es improbable que las conclusiones del MAR presenten un bajo riesgo de sesgo.