

# Lectura crítica de revisiones sistemáticas sobre estudios de prevención o tratamiento

Miguel Maldonado Fernández ■ Vicente Ruiz García

## OBJETIVOS DEL CAPÍTULO

- Definir las revisiones sistemáticas sobre intervenciones de prevención o tratamiento (ITP).
- Enumerar los elementos clave de validez de estas revisiones sistemáticas.
- Definir conceptos de riesgo de sesgo y de certeza en la evidencia de cada resultado.
- Describir la forma de realizar la síntesis cualitativa y cuantitativa.
- Reflexionar sobre la aplicabilidad de la evidencia extraída de las revisiones sistemáticas.
- Facilitar al lector la lectura crítica de una revisión sistemática tras la lectura de este capítulo.

## Introducción

La mejor evidencia sobre intervenciones de tratamiento o prevención (ITP) se obtiene de las revisiones sistemáticas (RS) de ensayos aleatorios. Estas revisiones recopilan de modo estructurado, consistente y prefijado toda la evidencia publicada o no sobre una pregunta PICO, extraen los datos de los estudios y construyen estimaciones agregadas de los efectos de las intervenciones (metaanálisis). Por ello son de enorme utilidad en la toma de decisiones sobre actividad clínica y en la elaboración de guías de práctica clínica, de modo que constituyen la piedra angular en el uso de la evidencia en la clínica. Cuando queremos saber si una intervención preventiva o una forma de tratar a los pacientes es mejor o peor que otra, por ejemplo, mejorar los reingresos o la mortalidad en una enfermedad u otra, el diseño ideal es una RS de ensayos clínicos que aborde la pregunta en cuestión. Fiarlo todo al resultado de un solo ensayo si existen muchos ensayos, no parece una respuesta sensata, pues podría proporcionarnos una respuesta que no represente la situación real y fuera más bien una situación particular. La gran cantidad de ensayos de ITP que se generan y la necesidad de estar al día hace que necesitemos documentos de síntesis que estén sujetos al menor sesgo posible y que representen una aproximación a «la verdad sobre la cuestión». El crecimiento de RS a lo largo de la última década ha sido exponencial, duplicando el número de las publicaciones cada 5 años. La proliferación de este tipo de RS hace que debamos conocer la forma en las que están construidas y presentadas, para decidir sobre la validez de las mismas.

## Escenario

Te encuentras en tu centro de salud pasando consulta. Acude una paciente de 17 años con un dolor de garganta de intensidad considerable, que le impide concentrarse para estudiar las pruebas de acceso a la Universidad, que tiene que hacer de forma inminente.

La paciente no tiene alergias medicamentosas ni otras patologías de interés. A la exploración no encuentras signos de absceso periamigdalino y únicamente evidencia una orofaringe eritematosa.

Su madre, muy preocupada por la situación de la paciente, te pide algún tratamiento que mitigue el dolor a la mayor brevedad, para que su hija pueda continuar con su estudio al máximo nivel. Una amiga íntima le ha comentado que su hija mejoró muchísimo cuando «le pincharon cortisona».

Afortunadamente, encuentras la siguiente revisión sistemática:

Corticosteroids for treatment of sore throat: systematic review and meta-analysis of randomised trials (1).

Leyéndolo con atención te preguntas:

- ¿Son útiles los corticoides para mejorar el dolor de garganta?
- ¿Darías un corticoide para mejorar el dolor de garganta en la opositora adolescente?

Sadeghirad B, et al. Corticosteroids for treatment of sore throat: systematic review and meta-analysis of randomised trials. *BMJ* 2017;358:j3887.

## Puntos clave para leer una RS

### APROXIMACIÓN A UNA DEFINICIÓN

Las RS sobre ITP son documentos que intentan reunir toda la evidencia empírica, publicada o no, que se ajuste a los criterios de elegibilidad especificados previamente, con el fin de responder a una pregunta de investigación concreta orientada al tratamiento y la prevención. Usan métodos explícitos y sistemáticos con el fin de minimizar sesgos, por lo tanto, proporcionan resultados más fiables que los de estudios individuales, y a partir de ellas se pueden extraer conclusiones y decisiones. Las características clave de una revisión sistemática son (1):

- Una pregunta clínica bien definida con formato PICO, con criterios de inclusión y exclusión predefinidos para los estudios.
- Una metodología explícita y reproducible.
- Una búsqueda sistemática que trata de identificar todos los estudios.
- Una evaluación de la validez de las conclusiones de los estudios incluidos, por ejemplo, a través de la evaluación del riesgo de sesgo entre los ensayos y de la calidad o certeza de la evidencia en cada resultado.
- Una presentación sistemática y una síntesis, que incluya las características y los hallazgos de los estudios incluidos.

### Pregunta clínica

Habitualmente, la RS de ITP se construye para responder a una pregunta clínica claramente formulada. La mejor forma de hacerlo es mediante el acrónimo PICO ya comentado al abordar el ensayo clínico en el capítulo 5.

**P (PARTICIPANTES).** A diferencia de lo que ocurre en los ensayos, donde los participantes son los pacientes o individuos, los participantes aquí son los propios ensayos clínicos (EC) realizados sobre pacientes que presentan una determinada condición objeto de nuestro interés. A diferencia del PICO de los ensayos, la incorporación de poblaciones diferentes a través de los estudios (edades, comorbilidad, diferentes estadios evolutivos, etc.) introduce una fuente de heterogeneidad que no aparece en los ensayos. Normalmente, en los textos se referencia que la unidad de análisis son los estudios. Solo en los metaanálisis de pacientes individuales la unidad de análisis son los enfermos. Estos se confeccionan con los datos crudos de los ensayos a los que acceden los investigadores, lo que permite explorar más información relevante.

**I (INTERVENCIÓN).** La intervención objeto de la revisión es la prevención o el tratamiento que se somete a prueba, ya sea un fármaco concreto, un grupo de ellos o cualquier otra actuación terapéutica, preventiva, ruta clínica u organizativa que ha sido estudiada en los ensayos originales.

**C (COMPARACIÓN).** La comparación es la acción alternativa con la que se contrasta la nueva intervención objeto del estudio. Esta puede ser un tratamiento falso o placebo, el tratamiento habitual de la enfermedad o incluso el mismo tratamiento en dosis diferentes o con vías de administración diferentes.

**O (OUTCOMES-RESULTADOS).** Los resultados son los eventos producidos objeto de la investigación. No todas las preguntas se contestan en una revisión y en ocasiones se necesita más de una revisión para contestar todas las preguntas relevantes. Los resultados deben ser lo más definidos posibles a través de la descripción de los métodos diagnósticos que se usaron, las escalas de medida, los umbrales a los que se ajustaron, el tiempo al que se midieron los resultados y las unidades de medida usadas. En el caso de emplear escalas hay que detallar los valores superiores e inferiores y si los valores más altos o los más bajos son los favorables.

Un asunto de importancia es lo ambigua o precisa que sea una pregunta clínica, pues las respuestas serán igualmente ambiguas o precisas. Esta cuestión varía con los criterios de inclusión de los participantes en los ensayos clínicos.

Un ejemplo de una pregunta con formato PICO sería: ¿En los enfermos con COVID-19 ingresados por bronconeumonía, el tratamiento con Remdesivir comparado con el placebo sirve para disminuir la mortalidad a 28 días o la estancia hospitalaria?

La amplitud de los componentes es crucial a la hora de extraer conclusiones, pues esta pregunta anterior podría haberse acotado aún más si incluyésemos solo a sujetos inmunocomprometidos o solo sujetos a partir de una determinada edad o sexo, dado que hoy sabemos que el riesgo de fallecer por COVID-19 si se sufre bronconeumonía puede depender de estos (2). En resumen, la pregunta más ambigua produce resultados globales que pueden no representar alguno de los subgrupos incluidos. Siguiendo el mismo ejemplo:

- P: Ensayos clínicos que incluyan enfermos con bronconeumonía y que estén ingresados en hospital.
- I: Remdesivir. La pregunta no incluye la dosis y esta podría ser diferente, dada a lo largo del día en solución continua o a una dosis fija administrada periódicamente.
- C: Placebo. Una solución lo más parecida a la intervención en aspecto y que se prescribe a las mismas horas y de la misma forma.
- O: Mortalidad a los 28 días y estancia hospitalaria. Pero podrían ser aún más, como efectos adversos, necesidad de oxígeno suplementario, carga viral, etc. Desde un punto de vista de importancia, los tres primeros son claramente resultados centrados en los pacientes, mientras que los dos últimos son prácticamente resultados subrogados que, aunque sean recogidos, no modificaran la incorporación o no del tratamiento si los resultados de importancia no producen beneficio. La RS de ITP deben estar centradas claramente en resultados centrados en los pacientes.

## Búsqueda de la evidencia y la exhaustividad

El lector debe revisar si hubo exhaustividad en la búsqueda. La búsqueda de los ensayos publicados se hace a través de bases de datos y la de los ensayos no publicados a través de los registros de ensayos clínicos, contacto con los autores y promotores y de la llamada literatura gris.

En una revisión sistemática deben incluirse los ensayos publicados y los no publicados. Si los autores no lo hicieran se podría incurrir en un sesgo de publicación.

La información publicada aparece recogida en bases de datos de artículos científicos. Idealmente se debería buscar al menos en Medline, Embase y CCRT (Cochrane Central Register of Trials). Existen numerosas bases de datos, muchas de ellas específicas de determinadas áreas de la salud: CancerLit, CiNAHL, WOS (Web of Science), PsycINFO, LILACS (Literatura Latinoamericana y del Caribe en Ciencias de la Salud), si se desea buscar literatura de esa zona concreta. Por ejemplo, al hacer una revisión sistemática de la enfermedad de Chagas, muy prevalente en His-

panoamérica, sería interesante que los autores incluyesen LILACS para buscar artículos de autores latinoamericanos. En cualquier caso, la elección de las bases de datos dependerá de los problemas clínicos que se aborden.

Los registros de EC como <http://controlled-trials.com>, <http://clinicaltrials.gov> y la World Health Organization (WHO) trial portal son bases de datos internacionales que constituyen una de las mayores fuentes de información sobre EC. Allí podemos encontrar: ensayos en sus diferentes fases sin haber concluido y ensayos terminados, en ocasiones con resultados sin publicar. Identificar estos registros facilita en futuras actualizaciones que no se pierdan estos si no se llegaron a publicar los resultados.

El lector debe revisar si los autores de la revisión sistemática se pusieron en contacto con los autores de los estudios individuales que se incluyeron en la revisión sistemática, ya que pueden ser fuente de información adicional de otros estudios no publicados, así como si hubo contacto con los patrocinadores de los estudios.

Asimismo, se debe chequear si se intentó encontrar en las referencias de los estudios localizados otros estudios originales que no estuvieran localizados. Por último, se debe revisar si se abarcó la literatura gris que abarca los repertorios de congresos, o tesis doctorales, que pueden ayudar a localizar pequeños ensayos clínicos que no se han llegado a publicar en revistas. La inclusión de literatura gris parece que incrementa solo discretamente los resultados a favor de la revisión a estudio (3).

## Selección de estudios y extracción de datos

El lector debe revisar cómo se realizó, pues como en el recuento de votos en unas elecciones, este proceso debe ser realizado por más de una persona. Los estudios que no cumplen los criterios de la pregunta PICO deben ser eliminados y detallarse la causa de su exclusión. A la hora de proceder a la extracción de datos debe revisarse si procedió de igual manera: si dos o más personas extrajeron los mismos. Lo normal es que este procedimiento, tanto la selección como la extracción, se haga de forma independiente. En realidad, vuelve a recordar el proceso de autenticación de las papeletas por los integrantes de la mesa electoral. Lo habitual son dos personas y si no existe consenso decide una tercera persona.

## Evaluación de la calidad de los ensayos

El resultado que se pueda extraer de una revisión sistemática depende, entre otros factores, de la validez de los artículos originales incluidos. Los puntos fundamentales sobre la validez de un ensayo clínico ya se citaron en el capítulo de ensayos.

La escala más utilizada para la detección de riesgo de sesgo es la de la Colaboración Cochrane (ya comentada en el capítulo 5), que es una evaluación basada en las características de los estudios: generación de la secuencia, ocultamiento de la secuencia, cegamiento de los participantes y del personal, cegamiento de los evaluadores de los resultados, datos de resultado incompletos, notificación selectiva de resultados, otras fuentes de sesgo. Cada característica se expresa como «bajo riesgo», «alto riesgo» o «riesgo poco claro» de sesgo (1).

Hoy sabemos que la interrupción precoz de los ensayos tiene más posibilidad de sesgo que los que se terminan sin interrupciones anticipadas. Interrumpir un estudio por seguridad, porque existen más muertes o efectos dañinos en uno de los grupos, es ético, pero interrumpir prematuramente por la aparición de un beneficio inesperado puede introducir sesgos importantes y debería tenerse en consideración (4).

## Resultados: metaanálisis y forest plot

El metaanálisis es un conjunto de procedimientos estadísticos que permite agregar los resultados de diferentes estudios entre sí. Si este procedimiento no se hace tras una revisión exhaustiva de acuerdo con lo que hemos descrito para una revisión sistemática, no tiene gran valor. El objetivo es mejorar el conocimiento al obtener un valor promedio de un valor desconocido que nos permita acercarnos al valor más real posible. Al incrementar el tamaño muestral se aumenta la precisión estadística.

Estos cálculos producen una media ponderada de los efectos de los tratamientos. Estos efectos dependen de los resultados estudiados. Si los valores son dicotómicos (por ejemplo, los pacientes mueren o no), los resultados los podemos presentar como diferencia absoluta entre vivos y muertos, también llamada diferencia absoluta de riesgos (DAR), proporción entre muertos y vivos, riesgo relativo u *odds ratio* (RR u OR), que ya se han explicado en el capítulo de EC. Si los valores son continuos, los resultados se presentan como diferencia de medias, bien de forma absoluta cuando están medidas en las mismas unidades, o de forma ponderada, cuando la forma de medirlas es a través de distintos instrumentos, como por ejemplo diferentes cambios en las escalas de depresión. El *forest plot* es el gráfico que presenta los resultados de cada estudio con el valor final agregado de todos los estudios.

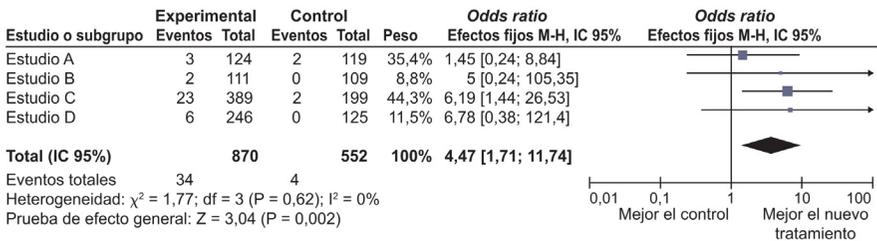
La estimación puntual de cada estudio individual se presenta como un cuadrado, cuyo tamaño es proporcional al peso de dicho estudio en el resultado combinado y la precisión de dicha estimación, es decir, los intervalos de confianza, se representan mediante una línea horizontal. En general los estudios de menor tamaño tienen menos precisión y su contribución a la estimación global del efecto recibe una menor ponderación.

El resultado combinado de todos los estudios se muestra como un rombo o diamante cuya amplitud horizontal representa los intervalos de confianza de la estimación global, es decir, el rango de valores entre los que el verdadero valor del efecto es probable que se sitúe. Su significación estadística se valora observando si su amplitud incluye la línea de no significación (1 si se trata de OR o RR y 0 si se trata de diferencia de riesgos o diferencia de medias). Cuando los resultados se presentan en forma de OR o RR el *forest plot* utiliza una escala logarítmica.

El *forest plot* permite explorar visualmente la existencia de heterogeneidad, viendo si las estimaciones puntuales de cada estudio son muy diferentes entre sí y en qué grado se solapan los IC. Hay diferentes estadísticos que miden heterogeneidad, pero el más frecuentemente usado es el  $I^2$ . Un valor de menos de 40% se considera indicativo de heterogeneidad baja. Entre 30-60% sugiere una probabilidad moderada de heterogeneidad, entre 50-90% puede ser sustancial y entre 75 y 100% considerable. Estos valores se solapan reafirmando la idea de que los intervalos se han establecido de manera orientativa.

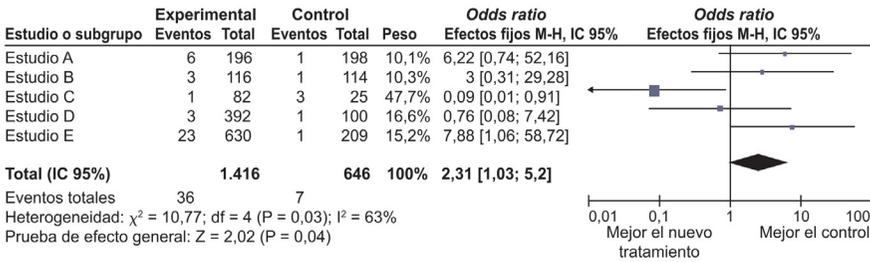
Grandes diferencias en las estimaciones puntuales de cada estudio e IC que no se solapan sugieren heterogeneidad. En cualquier caso, la heterogeneidad es una cuestión del clínico más que del estadístico. Las figuras 12.1 y 12.2 representan respectivamente un *forest plot* sin heterogeneidad y con heterogeneidad estadística. En la primera, los intervalos de confianza se solapan, mientras que en la segunda no todos los estudios se solapan y algunos claramente tienen resultados opuestos (C y E). El estadístico de heterogeneidad es 0 en el primero mientras que llega al 68% en el segundo.

Las diferencias que existen en los ensayos siempre van a introducir un grado de heterogeneidad en las revisiones: diferencias en fármacos, dosis, vías de administración, pacientes, estadio evolutivo, seguimientos, calidad de los estudios, etc. Sin la plausibilidad de que los tratamientos afecten de forma similar, a pesar de estas diferencias, no se deberían combinar los resultados individuales para extraer una conclusión global. Sin embargo, puede ser útil combinarlos para comprobar cómo afectan



**Figura 12.1** *Forest plot* que mide el riesgo de mejorar en la cefalea con el nuevo tratamiento en relación con el tratamiento control. El valor de  $I^2$  es de cero, no sugiriendo heterogeneidad estadística.

© Elsevier. Fotocopiar sin autorización es un delito.



**Figura 12.2** Forest plot que compara la mortalidad de un nuevo tratamiento para la cefalea. Los IC 95% de todos los estudios no se solapan y la  $I^2$  es del 63%, sugiriendo heterogeneidad importante.

estas diferencias al resultado final. Podría incluso considerarse hacer recomendaciones específicas para diferentes subgrupos de pacientes, grupos o tratamientos.

Los *forest plot* se realizan de acuerdo con dos asunciones: la del modelo de efectos fijos y la del modelo de efectos aleatorios. El primero asume ausencia de diferencias entre los estudios por poblaciones e intervenciones muy homogéneas, y el segundo, que estas diferencias pueden ser importantes. El segundo es más conservador y produce intervalos de confianza más amplios. Aunque en los protocolos se debe establecer qué modelo *a priori* se va a usar, lo habitual es confirmar la consistencia de los resultados para valorar la heterogeneidad con los dos modelos.

## ANÁLISIS DE SUBGRUPOS

El análisis de subgrupos o estratificado consiste en realizar metaanálisis para diferentes características o categorías de los estudios y valorar si los resultados combinados son diferentes entre sí (por ejemplo, si hay diferencias del efecto entre adultos y niños, entre dosis diferentes, lugares geográficos o calidad de la evidencia). Para comprobar esto y descartar que esta variación se deba al azar es necesario realizar la prueba de interacción, cuya significación inferior  $< 0,10$  sugiere un diferente efecto de la intervención en los subgrupos. Hay que recordar que los análisis de subgrupos son de tipo observacional y, salvo que tengamos datos individuales de los pacientes dentro de los ensayos, pueden producir hallazgos que en realidad no existan, pues los subgrupos pueden no estar balanceados al no derivarse de una asignación al azar. En el ensayo ISIS-2 se demostró beneficio de la aspirina sobre el placebo en todos los sujetos, pero en dos signos astrológicos de los pacientes, este no parecía existir (5). A más análisis de subgrupos, más posibilidad de producir hallazgos, que en realidad pueden no existir.

La metarregresión es una extensión del análisis por subgrupos que permite investigar simultáneamente el efecto de características categóricas (vía de administración, sexo, calidad de los estudios) y continuas (edad, dosis, tamaño muestral).

## ANÁLISIS DE SENSIBILIDAD

El análisis de sensibilidad pretende medir la consistencia del estimador final. Para ello se repite tantas veces el metaanálisis como estudios tenga la revisión, eliminándose en cada paso uno de ellos y dejando el resto. Si los resultados de los distintos metaanálisis son similares en magnitud, dirección y efecto, se puede concluir que los resultados son confiables o robustos. En caso contrario no se tendría un estimador robusto, lo cual exigiría cierta precaución en la interpretación de los resultados o podría ser motivo para generar nuevas hipótesis.

El análisis de sensibilidad también puede utilizarse para estudiar la influencia en los resultados del metaanálisis de ciertos aspectos relacionados con la validez de los estudios, como por ejemplo

la exclusión de estudios que no superan un determinado umbral de calidad, trabajos no publicados; o bien estudios en los que hay incertidumbre sobre el cumplimiento de los criterios de selección.

En ocasiones este análisis se confunde con el análisis de subgrupos. Los dos métodos difieren en dos aspectos. En primer lugar, los análisis de sensibilidad no intentan calcular el efecto de la intervención en el grupo de estudios eliminados del análisis, mientras que en los análisis por subgrupos se producen estimaciones para cada subgrupo. En segundo lugar, en el análisis de sensibilidad se hacen comparaciones informales entre diferentes formas de calcular lo mismo, mientras que en el de subgrupos se hacen comparaciones estadísticas formales entre los subgrupos.

## Certidumbre de la evidencia en los resultados

La calidad de la evidencia o certidumbre depende de dos factores. Por una parte, el riesgo de sesgo en cada uno de los estudios incluidos en la revisión. Por otro, el riesgo de sesgo de la propia revisión, debido a factores como la imprecisión, los sesgos de publicación, la evidencia indirecta, etc.

GRADE en las RS de ITP funcionaría como un ascensor que, instalado en los pisos altos (por tratarse de revisiones de ensayos clínicos), solo puede bajar la certidumbre de los resultados si se ve penalizado en alguna de las siguientes áreas:

### IMPRECISIÓN

Los estudios con pocos pacientes probablemente tendrán intervalos de confianza grandes, sobre todo si los resultados son poco numerosos. En este caso se recomienda bajar un escalón por imprecisión. Por regla general se sugiere bajar un escalón si los intervalos de confianza sobrepasan el 0,75 por abajo o el 1,25 por arriba.

Otra forma de medir imprecisión, además de los intervalos de confianza, es considerar si el número total de pacientes incluido en la revisión para el resultado estimado es inferior al del tamaño de un ensayo clínico con un tamaño muestral con suficiente poder estadístico. En este caso habría que bajar también un escalón por imprecisión.

Para estudios con pocos eventos y tamaños muestrales grandes en los grupos podría parecer que los riesgos relativos son similares. En este caso se aconseja usar medidas absolutas como las diferencias de riesgo y disminuir la confianza por imprecisión, si esta es amplia.

### INCONSISTENCIA

La inconsistencia se evalúa y se cuantifica a la baja cuando los valores promedios de los estudios individuales son claramente diferentes entre sí, sus intervalos de confianza no se solapan o los estadísticos que miden la heterogeneidad general son menores de  $p < 0,1$ . Si se ha hecho análisis de subgrupos, los análisis son creíbles y se muestran, no hay que bajar la calidad por inconsistencia. En caso de que los subgrupos sean plausibles, pero los resultados sean muy improbables y se exprese solo una estimación final de un solo valor final agregado, debería bajar la calidad al menos un escalón.

### NATURALEZA INDIRECTA

Cuando existen comparaciones indirectas de un fármaco A contra placebo y de B contra placebo, pero no existe comparación de A contra B, estamos ante una comparación indirecta. Otras formas de comparación indirecta incluyen diferencias en el tipo de población, por ejemplo, diferentes grados de enfermedad o diferentes cointervenciones, diferencias en la forma de medir el resultado o la existencia de un resultado subgrupo. En estos casos debería bajarse la calidad.

## RIESGO DE SESGO DE PUBLICACIÓN

Cuando los resultados son contrarios a la hipótesis de trabajo e incluyen pocos sujetos, es más probable que no se publiquen. Lo mismo sucede si el promotor es la industria farmacéutica. También se sabe que los autores, editores y revisores de revistas no son proclives a publicar los resultados, o los retrasan porque no los consideran prioritarios. Se aconseja disminuir un escalón la calidad si los estudios son pequeños, están financiados por la industria farmacéutica y los autores tienen claro conflicto de intereses con esta.

Cuando los autores o patrocinadores deciden no publicar cierto material del ensayo por la magnitud del resultado, la dirección de este o por la ausencia de significación estadística, se produce un sesgo conocido como sesgo de reporte (*reporting bias*) (6). Cuando es todo el ensayo el que no se publica, se produce un sesgo denominado sesgo de publicación o *publication bias*. Ejemplos de sesgo de publicación se han producido con antidepresivos (7) o antigripales (8). Organizaciones como *alltrials.net* están orientadas a que toda la investigación que se produzca sobre un campo se haga pública con la inclusión de todos los resultados, para minimizar el impacto de este sesgo. A pesar de que existe un mandato legal desde 2007 que en EE. UU. obliga a publicar los resultados de los ensayos incluidos en *clinicaltrials.gov*, en marzo de 2020 *alltrials.net* publicó: «Un juez de Nueva York ha dictaminado que cientos de ensayos clínicos registrados en *clinicaltrials.gov* están infringiendo la ley al no informar los resultados» El período analizado incluye 2007-2017 y todo parece indicar que las cosas no han cambiado mucho a este respecto (<https://www.alltrials.net/news/hundreds-of-clinical-trials-ruled-to-be-breaking-the-law/>).

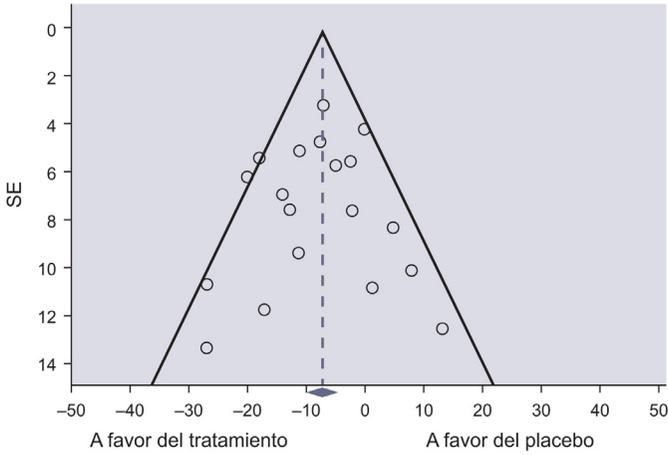
Existen diferentes formas de sugerir la presencia de sesgo de publicación a partir de gráficos y de pruebas estadísticas. El gráfico de embudo, *funnel plot*, es una de las formas más usadas. Explora la posibilidad de sesgo de publicación relacionando la magnitud del efecto combinado del tratamiento de cada estudio con la precisión. Cuando el diagrama de puntos de los diferentes estudios tiene una distribución simétrica alrededor del efecto, se sugiere la ausencia de sesgo de publicación. Y en su defecto indica una posible existencia de sesgo. La primera imagen de la [figura 12.3](#) muestra una distribución simétrica de los ensayos, mientras que la segunda imagen muestra una distribución con una asimetría en las bases que sugiere sesgo de publicación. Desgraciadamente, con menos de 10 estudios la posibilidad de sugerir sesgo de publicación a partir del *funnel plot* es baja y en realidad en la mayoría de las ocasiones no se presenta en las publicaciones (1).

La sospecha o confirmación de datos falsificados penaliza nuevamente este ítem. En el caso de apixabán, parece que los resultados de los metaanálisis publicados incluían aportaciones de un estudio que se demostró falsificado y su anulación en los metaanálisis producía cambios importantes en sus resultados en un tercio de estos (9).

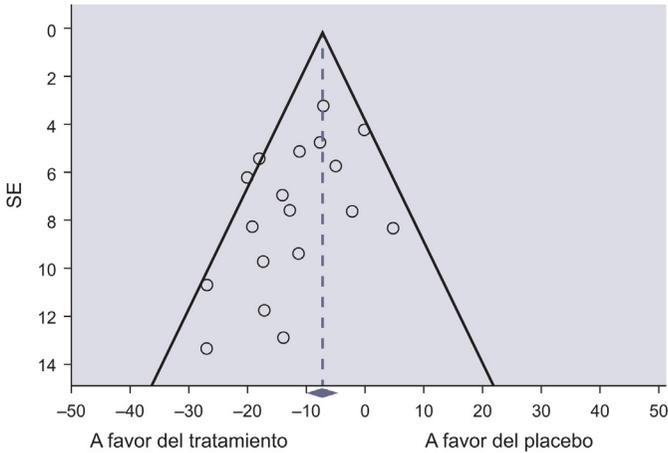
El sesgo de publicación nunca se puede descartar del todo, pero se intenta minimizar al explorar los metarregistros de ensayos clínicos, e idealmente al menos *clinicaltrials.gov* y el de la World Health Organization (WHO).

## Presentación de resultados: Summary of Findings (SoF) y perfil de evidencia

Es habitual encontrar en la parte final de los documentos un resumen estructurado con los hallazgos de la revisión, denominado SoF, Summary of Findings (resumen de hallazgos), o un perfil de evidencia de GRADE, que vamos a repasar en la [figura 12.4](#). Hemos adaptado los resultados del ensayo que usamos en este capítulo como ejemplo para elaborar un perfil de evidencia de acuerdo con GRADE. Las diferencias entre ambos son mínimas, pero en los dos el encabezado debe incluir los autores, la pregunta de investigación y el tipo de pacientes que incluye. En ambos se describen los resultados de interés, donde el lector debe comprobar que aparecen los resultados de carácter CRÍTICO o IMPORTANTE para resolver la pregunta en



A



B

**Figura 12.3** *Funnel plot* de una revisión sistemática. **A.** La distribución de los estudios a ambas partes del verdadero efecto sugiere que no existe sesgo de publicación, pues los estudios de menor tamaño muestral por azar se supone que se distribuirían a ambos lados del gráfico de forma parecida. **B.** La falta de estudios a la derecha de la línea media sugiere un posible sesgo de publicación.

cuestión. Se deberían dejar fuera de este resumen de evidencia los resultados subrogados y no centrados en los valores que resultan importantes para los pacientes. Idealmente no se deberían presentar más de siete resultados. Se deben mostrar los efectos secundarios, si los hubiera, para tener un balance riesgo/beneficio. En la tabla se expresan las unidades en que se mide cada resultado y la interpretación de estas, así como el tiempo de seguimiento en cada uno de estos.

**Autor(es):** Sadeghirad B, Siemieniuk RAC, Brignardello-Petersen R, Papola D, Lytvyn L, Vandvik PO, Merglen A, Guyatt GH, Agoritsas T.

**Pregunta:** Sin corticoides comparado con corticoides (ITT) para dolor de garganta

**Configuración:** Pacientes ambulatorios y de urgencias

**Bibliografía:** Corticosteroids for treatment of sore throat: systematic review and meta-analysis of randomised trials. BMJ 2017;358:j3887.

Número ensayos	Riesgo de sesgo	Inconsistencia	Evidencia indirecta	Imprecisión	Otras consideraciones	N Corticoides	N Placebo	RR (IC 95%)	MD (IC 95%)	Certidumbre	Importancia
<b>Resolución completa del dolor a las 24h</b>											
5	Bajo	Seria	Baja	Ninguna	Bajo	119/439 (27%)	67/424 (16%)	2,24 (1,17-4,29)	<b>124 más por 1.000</b> (17-129)	Moderada	Importante
<b>Prescripción de antibióticos</b>											
1	Bajo	Baja	Baja	Muy seria	Bajo			0,83 (0,61-1,13)	<b>96 menos por 1.000</b> (220 menos a 73 más)	Baja	Importante
<b>Tiempo medio hasta la resolución completa del dolor</b>											
6	Bajo	Seria	Baja	Seria	Bajo	44 h/353	33 h/367		11,1 h (21,8 menos a 0,4 menos)	Baja	Importante

**Figura 12.4** Perfil de evidencia de GRADE (adaptado). IC, intervalo de confianza; MD, diferencia media; RR, riesgo relativo.

**CUADRO 12.1 ■ Interpretación de calidad a través de GRADE**

**ALTA CALIDAD:** es muy improbable que las investigaciones adicionales modifiquen la certidumbre con respecto al cálculo del tamaño del efecto.

**MODERADA:** probablemente las investigaciones futuras tendrán una **repercusión importante** en la certeza respecto al cálculo del efecto y pueden modificarlo.

**BAJA:** muy probablemente investigaciones adicionales tendrán una repercusión importante en la certeza respecto al cálculo del efecto y **es muy posible que lo modifiquen**.

**MUY BAJA:** cualquier cálculo del efecto es muy poco fiable. Es muy necesaria más investigación para estimar el efecto.

Se debe describir si existe riesgo de sesgo en cada una de las áreas de GRADE, etiquetándolo como Ninguno/no; Serio/Muy serio. Si el resultado es cualquiera de los dos últimos, hay que poner una llamada para explicar por qué se ha etiquetado de esta forma. Por último, se expresan los resultados de los estudios, que incluyen: pacientes incorporados en cada rama para cada resultado, número de ensayos incluidos para cada resultado, y el estimador relativo y/o absoluto. Finalmente, la certeza que supone ese resultado y, en caso de los SoF, una interpretación literal de la misma. La gradación va desde una certeza muy importante en los efectos que nos presentan a una muy baja certeza en la estimación de estos (cuadro 12.1).

En el caso del SoF del artículo que es analizado en este capítulo, para el primer resultado (resolución completa del dolor a las 24 h), se etiqueta como certidumbre **MODERADA** y se describe como: «los corticoides *probablemente* incrementan la posibilidad de una resolución completa del dolor a las 24 h», mientras que el resultado de «la prescripción de antibióticos» o «tiempo medio hasta la resolución completa del dolor» son etiquetados como certidumbre **BAJA** y se describen de la siguiente manera: «los corticoides *pueden* disminuir la posibilidad de tomar antibióticos o *pueden* acortar la duración del dolor».

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) es el consenso de escritura para desarrollar una revisión sistemática en la actualidad. Seguir cada uno de los apartados permite evaluar todos los apartados de una revisión sistemática (10).

Es habitual que los protocolos de las revisiones estén disponibles en registros como PROSPERO, en la Universidad de York (<https://www.crd.york.ac.uk/PROSPERO/>) o en publicaciones como en la Colaboración Cochrane.

## Aplicabilidad de los tratamientos

### DISCUSIÓN

En la parte final de una RS, tras exponer los resultados, el lector debería encontrar una reflexión en la que se hiciera un balance sobre lo completa o no que es la misma. No es infrecuente que los tiempos de seguimiento para determinadas variables que son importantes en la revisión sean cortos o los efectos secundarios se hayan medido en pocos ensayos.

Otra de las cuestiones para revisar es la certeza sobre los efectos de las intervenciones que se ha examinado, pues en los etiquetados como bajos o muy bajos, cualquier nuevo estudio podría cambiar el sentido del efecto, o en el último caso el estimador puede ser cualquier otro distinto al reportado.

Debiera tomarse en consideración si existen comparaciones solo contra placebo y se han obviado comparaciones contra otras alternativas igualmente eficaces. El lector debe revisar si se plantea un balance riesgo/beneficio entre la mejora que produce el tratamiento que se pone a prueba y los posibles efectos adversos. Así, otros actores emergen a la hora de decidir más allá de la calidad de la evidencia: las preferencias de los pacientes y los efectos adversos (11).

## EVIDENCIA EN CONTEXTO

Poner a la evidencia reportada en contexto con otras revisiones sistemáticas producidas sobre el mismo tema permite comprobar si existen diferencias importantes sobre las conclusiones acerca de la misma. Diferencias mínimas en la formulación de la pregunta de investigación pueden cambiar los criterios de inclusión de los estudios, agregar diferentes comparadores o tiempo de seguimiento y producir resultados distintos.

## CONSECUENCIAS PARA LA INVESTIGACIÓN

Con frecuencia se señalan reflexiones acerca de las consecuencias que la RS tendrá para la investigación futura, pues se pueden poner en evidencia claras lagunas al respecto, por falta de estudios que incluyan a pacientes en un estadio concreto de enfermedad, duración insuficiente del seguimiento o falta de abordaje de resultados en salud necesarios para tomar una decisión bien informada, como pueden ser calidad de vida, efectos secundarios u otros.

## VISIÓN GLOBAL FRENTE A LA VISIÓN PARTICULAR DE LOS SUBGRUPOS (*LUMPERS VS. SPLITTERS*)

Hay que decidir hasta qué punto esa RS puede aplicarse a pacientes (12) y entornos concretos (13). Aunque el conocimiento que se genere pueda ser válido a nivel general, la toma de decisión en pacientes concretos exige una aplicación local o individual (globalizar la evidencia pero localizar las decisiones) (14).

El lector debe ser consciente de hasta qué punto las recomendaciones generales se pueden adaptar a pacientes específicos o no. Imaginemos una estrategia para prevenir el cáncer de colon que se realiza en toda España. Los diferentes ensayos en cada provincia demuestran que es efectivo en todas las provincias menos en Valencia, y sus pacientes son de esta provincia. ¿Debiera dejar de recomendar esta estrategia porque a nivel local no funcionó? La respuesta parece obvia, la información recopilada de la mayoría de los estudios sugiere que debe funcionar y deberíamos aceptar estos resultados, salvo que en esta provincia existan características específicas que la hagan bien diferente del resto. En otras ocasiones los valores de todos los estudios «no dejan ver el bosque» y la asunción de los *lumpers* en relación con los *splitters* no está tan clara. En la RS sobre intervenciones para mejorar la adherencia a la medicación para la insuficiencia cardíaca (15,16) se sugiere que las estrategias son efectivas para la misma. Sin embargo, la RS incluye multitud de intervenciones y no diferencia entre insuficiencia cardíaca con baja fracción de eyección y fracción de eyección conservada. Se sugiere una reducción de mortalidad y de reingresos, pero nos haría falta mucha más información para saber qué intervención de las múltiples que señala es efectiva y para qué tipo de insuficiencia cardíaca. Esta información desagregada (*splitter*) sería de gran utilidad.

Un comentario final: Los clínicos debemos hacer el esfuerzo de considerar si la evidencia centrada en resultados clínicamente relevantes para los pacientes está suficientemente bien construida o no. Los pacientes reúnen características únicas, pertenecen de diferentes ámbitos, tienen diferentes valores y comorbilidad y una percepción del riesgo distinta. La pericia de un buen clínico es aunar todas ellas.

## Artículo

Sadeghirad B, et al. Corticosteroids for treatment of sore throat: systematic review and meta-analysis of randomised trials. *BMJ* 2017;358:j3887. Disponible en: <http://dx.doi.org/10.1136/bmj.j3887>.

## Plantilla CASPe contestada para este artículo concreto

En el cuadro 12.2 se muestra la plantilla CASPe contestada para este artículo concreto.

**CUADRO 12.2 ■ Evaluación crítica del artículo propuesto (plantillas CASPe)**

**A) ¿Los resultados de la revisión son válidos?**

**Preguntas «de eliminación»**

	Sí ✓	No sé	No
<p><b>1. ¿Se hizo la revisión sobre un tema claramente definido?</b></p> <p><i>PISTA: una pregunta clínica se formula con los componentes PICO:</i></p> <ul style="list-style-type: none"> <li>• La población de estudio.</li> <li>• La intervención realizada.</li> <li>• La comparación.</li> <li>• Los desenlaces (outcomes) considerados.</li> </ul>	Sí ✓	No sé	No
<p>Sí. La revisión se hizo sobre un tema que puede traducirse en una pregunta PICO (Paciente, Intervención, Control, <i>Outcome</i> o resultado):</p> <p><b>Pacientes:</b> publicaciones de ensayos que incluyen pacientes de 5 o más años atendidos en urgencias con faringitis, amigdalitis o síndrome de dolor de garganta. Por dolor de garganta incluye un abanico de descripciones que van desde el dolor en sí a la faringitis o el dolor al tragar. Los criterios de exclusión los define claramente e incluyen: inmunocomprometidos, mononucleosis, dolor de garganta poscirugía o postintubación, reflujo, etc.</p> <p><b>Intervención:</b> corticoides. No describe la dosis, ni la vía de administración, aunque en la introducción se menciona que son dosis bajas a moderadas.</p> <p><b>Control:</b> tratamiento habitual (<i>standard of care</i>) o placebo.</p> <p><b>Outcomes:</b> resolución completa del dolor a las 24 y 48 h; tiempo medio de inicio de la reducción del dolor; tiempo medio hasta la resolución completa del dolor; resolución absoluta del dolor a las 24 h; duración de los síntomas «intolerables»; recidiva del dolor; días perdidos de trabajo o escuela; necesidad de antibióticos; y efectos adversos de la intervención.</p>			

	Sí ✓	No sé	No
<p><b>2. ¿Buscaron los autores el tipo de artículos adecuado?</b></p> <p><i>PISTA: el mejor «tipo de estudio» es el que:</i></p> <ul style="list-style-type: none"> <li>• Se dirige a la pregunta objeto de la revisión.</li> <li>• Tiene un diseño apropiado para la pregunta (puede ser óptimo o subóptimo por razones diversas).</li> </ul>	Sí ✓	No sé	No
<p>Sí. Se ha hecho una revisión sistemática de ensayos clínicos aleatorizados, que es el tipo de estudios que responde a preguntas sobre prevención o tratamiento.</p>			

**¿Merece la pena continuar?**

(Continúa)

Preguntas detalladas			
<b>3. ¿Crees que estaban incluidos los estudios importantes y pertinentes?</b>	Sí ✓	No sé	No
<p><i>PISTAS 1:</i></p> <ul style="list-style-type: none"> <li>• ¿Qué bases de datos bibliográficos se han usado? ¿Qué estrategia de búsqueda?</li> <li>• Seguimiento de las referencias.</li> <li>• Contacto personal con autores.</li> <li>• Búsqueda de estudios no publicados y de resultados no publicados-control de los protocolos originales.</li> <li>• Idiomas distintos del inglés.</li> </ul> <p><i>PISTAS 2:</i></p> <ul style="list-style-type: none"> <li>• Criterios de inclusión/exclusión.</li> <li>• Selección de estudios.</li> <li>• Extracción de datos.</li> </ul>	<p>Sí. Esta revisión es una actualización de otra previa hecha para Cochrane. Aceptan por buena la búsqueda bibliográfica realizada en esa primera revisión, y actualizan la búsqueda desde el 1 de enero de 2010. De ese modo se solapan 2 meses con lo buscado en la revisión Cochrane. La búsqueda se llevó a cabo en Medline, Embase y en el Cochrane Central Register of Controlled Trials (CENTRAL). Buscaron también en clinicaltrials.org, pero no en metarregistros o en la plataforma de registros internacionales de ensayos clínicos (ICTRP) de la OMS. No hubo restricción por idioma (es decir, no se rechazaron ensayos clínicos que no estuviesen en inglés). No mencionaron la búsqueda de literatura gris. Tampoco se hizo referencia a si se pusieron en contacto con los autores para recabar información e indagar sobre otros estudios no publicados. Sí buscaron en las referencias bibliográficas de los ensayos clínicos incluidos.</p>		
<b>4. ¿Crees que los autores de la revisión han hecho suficiente esfuerzo para valorar la calidad de los estudios y de los resultados incluidos?</b>	Sí ✓	No sé	No
<p><i>PISTA 1: el «riesgo de sesgo» depende (Cochrane RoBI):</i></p> <ul style="list-style-type: none"> <li>• Secuencia RZ.</li> <li>• Ocultación de la secuencia de aleatorización.</li> <li>• Cegamiento.</li> <li>• Pérdidas (atrición).</li> <li>• Reporte selectivo de desenlaces.</li> <li>• Otros.</li> </ul> <p><i>PISTA 2: valorar para cada desenlace, en cada estudio y en el conjunto de estudios.</i></p>	<p>Sí. Cuatro revisores trabajando por pares extrajeron los datos y evaluaron de forma independiente (es decir, cada uno trabajando por su lado, no revisando a la vez los artículos) la calidad metodológica de cada ensayo con la herramienta de medición del riesgo (Risk of bias tool: RoB) de Cochrane, que evalúa aleatorización, cegamiento, reporting bias y otros riesgos. También evaluaron la certidumbre en cada resultado a través de GRADE. Seleccionan los estudios de acuerdo a criterios de inclusión y exclusión definidos en los que añadieron un apartado para la exclusión de menos de 5 años por lo difícil de valorar el dolor de forma creíble a esa edad. De forma duplicada extrajeron los datos de acuerdo a: autores, año, localización, diagnóstico, tamaño de estudio, edad, etc.</p> <p>En el apéndice 2 se describe el riesgo de sesgo para cada uno de los estudios. Diez ensayos en los que en cuatro hay algún dominio con alto riesgo de sesgo, especialmente en el ensayo de Ahn. La figura 2 estudia la resolución completa del dolor a las 24 h. RR 2,24 (1,17-4,29). El dolor es una variable subjetiva que bien podría verse afectada por la ausencia de cegamiento. Revisemos el riesgo de sesgo de los ensayos para este resultado de los estudios involucrados. La descripción de estos está en el apéndice 2 como tabla adicional. La descripción que se ve es alto riesgo de sesgo por falta de ocultación de la secuencia de la aleatorización en los artículos de Ahn y Marvez Valls y falta de cegamiento del resultado en Ahn y O'Brien. El resto está de color verde, lo que implica bajo riesgo de sesgo. Ninguno de estos ensayos aparece para este resultado, lo que significa que la valoración que debiera hacerse es bajo riesgo de sesgo.</p>		

5. Si los resultados de los diferentes estudios han sido mezclados para obtener un resultado «combinado», ¿era razonable hacer eso?	Sí ✓	No sé	No
<p>PISTA: la heterogeneidad puede ser:</p> <ul style="list-style-type: none"> <li>• Clínica.</li> <li>• Metodológica.</li> <li>• Estadística (<math>I^2</math>).</li> </ul>	<p>Es decir, ¿son los ensayos clínicos suficientemente homogéneos para combinarlos entre sí?</p> <p>Hay una heterogeneidad clínica manifiesta que se tiene en cuenta <i>a priori</i> debida a varios factores: edad (niños-adultos), postulando un mayor efecto en adultos; ruta de administración (oral-parenteral), postulando un mayor efecto en la parenteral; cultivos positivos a estreptococo beta hemolítico A, postulando un mayor efecto en los que tenían cultivo positivo; lugar de la atención médica (atención primaria/urgencias), postulando un mayor efecto en los pacientes atendidos en urgencias; y lugar del tratamiento posterior (hospitalización-tratamiento ambulatorio).</p>		
<b>B) ¿Cuáles son los resultados?</b>			
<p>6. ¿Cuál es el resultado global de la revisión?</p> <p>PISTA: considera:</p> <ul style="list-style-type: none"> <li>• Valora para los desenlaces positivos y también los negativos.</li> <li>• ¿Cuáles son los resultados para cada desenlace?</li> <li>• ¿Cómo están expresados los resultados? RR, HR, etc.</li> <li>• ¿Muestran gráficos forest plot?</li> </ul>	<p>Diez estudios de los cuales tres son en niños, uno en adultos y niños y el resto solo en adultos. Está recogida en una tabla resumen (SoF) donde se clarifican los resultados prefijados para cada uno de ellos.</p> <p>Ver tabla 1 del artículo.</p> <p>Resolución completa del dolor a las 24 h RR 2,24 (1,17-4,29). Resolución del dolor a las 48 h RR 1,48 (1,26-1,75). Reaparición de los síntomas RR 0,52 (0,16-1,73). Prescripción de antibióticos 0,83 (0,61-1,13). Reducción del dolor en 24 h MD 1,3 (0,7-1,9). Días perdidos de escuela. Los describe de forma individual en los dos ensayos, pero no da un valor agregado.</p> <p>Efectos secundarios severos. Los describe, pero no da un valor agregado y señala que son similares en ambos grupos.</p> <p>Se aprecian <i>forest plot</i> para diferentes resultados en las figuras 2, 3, 4, 5 y 6.</p> <p>La figura 2 del artículo representa la resolución completa del dolor a las 24 h y muestra cinco ensayos cada uno con un IC distinto que se solapan entre sí excepto dos (Tasar y Hayward). Cada estudio está representado por un cuadrado que es más o menos grande en función del peso en el resultado del agregado final. Los intervalos de confianza aparecen como una línea recta que atraviesa cada cuadrado y representa en sus extremos los valores inferiores y superiores del intervalo. Todos los valores medios de los ensayos son favorables a los corticoides, pero en tres de ellos, incluido el de mayor peso, no son estadísticamente significativos. Aparece un valor final agregado en forma de rombo que no incluye el 1, por lo que se considera como estadísticamente significativo. Un RR 2,24 (1,17-4,29) significa que aproximadamente es dos veces más probable que desaparezca el dolor de garganta si se toma corticoide que si no se toma, pero que este valor podría llegar a ser hasta cuatro veces más o apenas igual tomar placebo. El <i>forest plot</i> incluye un valor de <math>I^2</math> de heterogeneidad de 69%, que expresa una heterogeneidad importante.</p> <p>Se evalúan resultados positivos pero también efectos secundarios de forma genérica, en forma de efectos adversos severos.</p>		

(Continúa)

7. Para el conjunto de los estudios (en cada desenlace concreto)	Sí ✓	No sé	No
<ul style="list-style-type: none"> <li>• «Riesgo de sesgo en los desenlaces».</li> <li>• ¿Cuál es la precisión de los resultados?</li> <li>• ¿Son consistentes los resultados de los estudios para cada desenlace?</li> <li>• ¿Es indirecta la evidencia en algún desenlace?</li> <li>• Sesgo de publicación para ese desenlace incluido en el protocolo pero no mostrado.</li> </ul>	Sí.	<p><b>Riesgo de sesgo en los desenlaces:</b> para la resolución del dolor a las 24 h ya ha sido comentado con anterioridad.</p> <p><b>Inconsistencia:</b> el material auxiliar del apéndice 3 explora si existe interacción entre los subgrupos y analiza los supuestos antes comentados (adultos/niños, oral/parenteral, atención primaria/urgencias, etc.) y no la encuentra. Es evidente que existe heterogeneidad clínica, de tratamientos, etc., hay una I<sup>2</sup> en la figura de 69%, pero no se degrada debido a que se ha hecho análisis de subgrupos, los análisis son creíbles y se muestran en la información adicional. De hecho, los autores declaran que no bajan la calidad por inconsistencia.</p> <p><b>Evidencia indirecta:</b> hay diferencias en la población pues se incluye a niños al menos en esta variable, lo que implica bajar un escalón. También hay efectos indirectos derivados del hecho de que algunos de estos estudios para esta variable incluyen dexametasona (DXM) IM y otros DXM oral, diferentes dosis de DXM oral y otros corticoides como prednisona y betametasona, lo que debería llevar asociado bajar otro escalón al menos. GRADE aconseja bajar al menos dos escalones si hay más de una fuente de naturaleza indirecta afectada. Los autores no bajaron la calidad ningún escalón por este ítem.</p> <p><b>Imprecisión:</b> los autores bajan un escalón por imprecisión. Lo justifican porque el límite inferior sugiere beneficio discreto, y el límite superior, gran beneficio. Hemos calculado las diferencias absolutas, como aconseja GRADE, y son de 0,19 (0,04-0,33). No nos parecen tan imprecisos, ni los valores absolutos ni los relativos, por lo que nuestra valoración no se degradaría en este apartado.</p> <p><b>Sesgo de publicación:</b> no se reporta.</p> <p>Finalmente, por motivos distintos a los autores de la revisión, la calificación quedaría en MODERADA para este resultado. Esto sugiere que existe una certidumbre moderada para quedar sin dolor a las 24 h.</p>	
<b>C) ¿Son los resultados aplicables en tu medio?</b>			
8. ¿Se pueden aplicar los resultados en tu medio?	Sí ✓	No sé	No
<p><i>PISTA: considera si:</i></p> <ul style="list-style-type: none"> <li>• Los pacientes cubiertos por la revisión pueden ser suficientemente diferentes de los de tu área.</li> <li>• Tu medio parece ser muy diferente al del estudio.</li> </ul>	<p>La enferma podría estar representada por cualquiera de los ensayos que aparecen en la revisión. Los corticoides son los mismos, los cuidados son similares, la prevalencia de comorbilidad podría ser similar, etc. No hay nada diferente que nos sugiera que estamos alejados del escenario propuesto.</p> <p>Es más, los resultados incluyen la representación de estudios que obtienen pacientes de atención primaria, no solo de situaciones de mucha más severidad como podrían ser pacientes atendidos en urgencias.</p>		

9. ¿Se han considerado todos los resultados importantes para tomar la decisión?	Sí ✓	No sé	No
<ul style="list-style-type: none"> <li>• <i>Utilidades y disutilidades.</i></li> <li>• <i>Balance de efectos positivos/negativos.</i></li> <li>• <i>Preferencias del paciente, costes, etc.</i></li> </ul>	<p>Los efectos adversos solo se reportaron en tres estudios, pero no parecieron adjudicarse a los corticoides y estuvieron más en relación con la propia enfermedad. Como es habitual, los efectos adversos están infradiagnosticados, pero fueron similares. Los precios de los corticoides en el momento actual son bajos y no hay posibilidad de discriminación de posible uso. Incluso en enfermos diabéticos es poco probable que afecten más de 1 día los ajustes de insulina.</p> <p>Dentro del dolor, la percepción de beneficio clínico de los pacientes es muy importante. Hoy se sabe que este umbral para el dolor leve es de 6 (4-8 puntos)/100 VAS, de 13 (11-14)/100 y de 21 (20-23)/100 para los dolores moderados y severos respectivamente (17). La diferencia media de usar corticoides al cabo de 24 h es 1,3 (0,7-1,9)/10, lo que equivale a un valor de 13/100. Incluso para niveles de dolor moderado existirá beneficio clínico para la paciente de nuestro escenario.</p> <p>Existen evidencias de que los pacientes que sufren infecciones de vías altas (incluida la faringitis) ven mermada su calidad de vida, medida mediante la escala SF-36 (18). Los corticoides en dosis utilizada en el ejemplo (una dosis baja, de administración única) parece que no provocarían efectos secundarios muy significativos. Por lo tanto, es posible que sea buena idea tratar a nuestro paciente con corticoides.</p>		

**Bibliografía**

1. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd ed. Chichester: John Wiley & Sons; 2019.
2. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. *BMJ* 2020;370:m3339.
3. Hopewell S, McDonald S, Clarke MJ, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database Syst Rev* 2007;(2):MR000010.
4. Latour Pérez J, Cabello López JB. Interrupción precoz de los ensayos clínicos. ¿Demasiado bueno para ser cierto? *Med Int* 2007;31(9):518-20.
5. Sleight P. Debate: Subgroup analyses in clinical trials: fun to look at - but don't believe them! *Curr Control Trials Cardiovasc Med* 2000;1(1):25-7.
6. Dwan K, Gamble C, Williamson PR, Kirkham JJ. The Reporting Bias Group. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias — An Updated Review. *PLoS ONE* 2013;8(7):e66844.
7. Turner EH, Tell RA. Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy. *New Engl J Med* 2008;358:252-60.
8. Doshi P, Jefferson T, Del Mar C. The Imperative to Share Clinical Study Reports: Recommendations from the Tamiflu Experience. *PLoS Med* 2012;9(4):e1001201.
9. Garmendia CA, Nassar Gorra L, Rodriguez AL, Trepka MJ, Veledar E, Madhivanan P. Evaluation of the Inclusion of Studies Identified by the FDA as Having Falsified Data in the Results of Meta-analyses: The Example of the Apixaban Trials. *JAMA Intern Med* 2019;179(4):582-4.
10. PRISMA. [Fecha de última consulta: 20 de septiembre de 2020]. Disponible en: <http://prisma-statement.org/Extensions/>.
11. Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation — determinants of a recommendation's direction and strength. *J Clin Epidemiol* 2013;66:726-35.

© Elsevier. Fotocopiar sin autorización es un delito.

12. Murad MH, Montori VM, Ioannidis JPA, Jaeschke R, Devereaux PJ, Prasad K, et al. How to Read a Systematic Review and Meta-analysis and Apply the Results to Patient Care: Users' Guides to the Medical Literature. *JAMA* 2014;312(2):171-9.
13. Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation — determinants of a recommendation's direction and strength. *J Clin Epidemiol* 2013;66(7):726-35.
14. Eisenberg JM. Globalize The Evidence, Localize The Decision: Evidence-Based Medicine And International Diversity. *Health Affairs* 2002;21(3):166-8.
15. Khazanie P, Allen LA. Meta-Analysis of Medication Adherence Interventions Among Heart Failure Patients: The Lumpers Make Their Case. *J Am Heart Assoc* 2016;5(6):e003827.
16. Ruppert TM, Cooper PS, Mehr DR, Delgado JM, Dunbar-Jacob JM. Medication Adherence Interventions Improve Heart Failure Mortality and Readmission Rates: Systematic Review and Meta-Analysis of Controlled Trials. *Journal of the American Heart Association* 2016;5(6):e002606.
17. Olsen MF, Bjerre E, Hansen MD, Tendal B, Hilden J, Hróbjartsson A. Minimum clinically important differences in chronic pain vary considerably by baseline pain and methodological factors: systematic review of empirical studies. *J Clin Epidemiol* 2018;101:87-106. e2. Disponible en: <https://tinyurl.com/y4nmhgpt>.
18. Linder JA, Singer DE. Health-related quality of life of adults with upper respiratory tract infections. *J Gen Intern Med* 2003;18(10):802-7.

### **Cómo citar este capítulo:**

Maldonado M, Ruiz V. Lectura crítica de revisiones sistemáticas sobre estudios de prevención o tratamiento. En: Cabello Juan B, editor. *Lectura crítica de la evidencia clínica*, 2.ª ed. Barcelona: Elsevier; 2022. p. 135-152.