AMP Acta Médica Peruana

ARTÍCULO ESPECIAL

Aspectos básicos sobre la lectura de revisiones sistemáticas y la interpretación de meta-análisis

Some basic concepts about reading systematic reviews and interpreting meta-analysis

Jose E. Fernandez-Chinguel[1], Jessica H. Zafra-Tanaka[2], Sergio Goicochea-Lugo[3], Christopher I. Peralta[4], Alvaro Taype-Rondan[5]

1 Investigador independiente. Chiclayo, Perú.
2 CRONICAS Centro de Excelencia en Enfermedades Crónicas, Universidad Peruana Cayetano Heredia. Lima, Perú.
3 Instituto de Evaluación de Tecnologías en Salud e Investigación, EsSalud. Lima, Perú
4 Sociedad Científica de Estudiantes de Medicina Villarrealinos, Universidad Nacional Federico Villarreal. Lima, Perú.
5 Unidad de Investigación para la Generación y Síntesis de Evidencias en Salud, Universidad San Ignacio de Loyola. Lima, Perú.

Correspondencia
Alvaro Taype-Rondan
alvaro.taype.r@gmail.com

Recibido: 15/01/2019
Arbitrado por pares
Aprobado: 03/04/2019

Citar como: *Fernandez-Chinguel JE, Zafra-Tanaka JH, Goicochea-Lugo S, Peralta CI, Taype-Rondan A. Aspectos básicos sobre la lectura de revisiones sistemáticas y la interpretación de meta-análisis. Acta Med Peru. 2019;36(2):157-69*

RESUMEN

Las revisiones sistemáticas (RS) son estudios que buscan resumir la evidencia disponible sobre una pregunta de investigación, para lo cual pueden usar estrategias estadísticas conocidas como meta-análisis (MA). En la actualidad, las RS son fundamentales para tomar decisiones basadas en evidencias, por lo cual resulta de suma importancia que los profesionales de la salud sepan enfrentarse a este tipo de estudios. Por ello, el presente artículo tiene por objetivo familiarizar al lector con los aspectos básicos para realizar una correcta lectura de RS e interpretación de MA, para lo cual se utiliza un ejemplo hipotético de una condición ficticia llamada "Síndrome del glotón". Asimismo, se aborda la interpretación de la certeza de la evidencia según la metodología Grading of Recommendations Assessment, Development, and Evaluation (GRADE).

Palabras clave: Educación Médica; Educación Continua; Medicina basada en la evidencia; Metaanálisis (fuente: DeCS BIREME)

ABSTRACT

Systematic reviews (SR) are studies that seek to summarize the available evidence regarding a research question, for which they can use statistical strategies known as meta-analysis (MA). Currently, SRs are fundamental to making evidence-based decisions, which is why it is very important for health professionals to know how to face this type of studies. Therefore, this article aims to familiarize the reader with the basic concepts to make a correct appraisal of SRs and interpretation of MAs, for which a hypothetical example of a fictitious condition called "Glutton Syndrome" is used. In addition, the interpretation of the certainty of the evidence according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology is addressed.

Keywords: Education, medical; Education, continuing; Evidence-based medicine; Meta-analysis (source: MeSH NLM)

REVISIÓN SISTEMÁTICA

Las revisiones sistemáticas (RS) son estudios secundarios que buscan responder una pregunta de investigación para lo cual realizan búsquedas exhaustivas de la evidencia disponible (estudios que hayan respondido a dicha pregunta de investigación) y sintetizan los resultados encontrados en dichas investigaciones. Este procedimiento requiere de una metodología crítica, transparente y reproducible. De esta manera, las RS buscan proporcionar la mejor evidencia disponible -hasta la fecha en la que realiza sus búsquedas- sobre las preguntas que están abordando, por lo que son consideradas como el pilar para el proceso de toma de decisiones basadas en evidencia [1,2].

Las RS pueden contestar diversas preguntas de investigación. Por ejemplo, si se desea comparar dos o más intervenciones sanitarias, se formulará una pregunta que especifique: la población de estudio, la intervención a evaluar, con qué se comparará esta intervención y los desenlaces. Esto se conoce como pregunta PICO (de las palabras en inglés: *population, intervention, comparison, outcome*). Sin embargo, existen otras RS que pueden plantear otro tipo de preguntas de investigación. En la Tabla 1 se muestra una categorización de RS adaptada de la propuesta por Munn, en base al tipo de pregunta de investigación que aborda [1].

Para elaborar una RS se deben seguir sistemáticamente una serie de pasos que resumimos a continuación [2-5]:

Tabla 1. Tipos de revisiones sistemáticas, adaptado de Munn.

Tipo de RS	Formato de pregunta	Objetivo
Eficacia o daños de intervención	En pacientes con depresión mayor, ¿realizar ejercicios en comparación con no realizarlos mejora el estado depresivo? • *Población:* pacientes con depresión mayor • *Intervención:* realizar ejercicio • *Comparador:* no realizar ejercicio • *Desenlace:* mejora el estado depresivo	Evaluar la eficacia y los daños causados por una intervención. Esto permite elegir una intervención.
Etiología y / o riesgo	¿Los adultos que trabajan en una planta nuclear tienen mayor riesgo a desarrollar cáncer? • *Población:* adultos • *Exposición:* trabajar en una planta nuclear • *Control:* no trabajar en una planta nuclear • *Desenlace:* desarrollo de cáncer	Evaluar la relación entre una exposición y un resultado de salud, y en qué medida esto permite elegir una prueba diagnóstica.
Exactitud de pruebas diagnósticas	En mujeres con cáncer de mama, ¿cuál es la exactitud de la prueba diagnóstica de la tomografía comparada con la biopsia? • *Población:* pacientes con cáncer de mama • *Prueba diagnóstica:* tomografía • *Desenlace:* exactitud de la prueba	Evaluar la exactitud de las pruebas diagnósticas que identifican la presencia o ausencia de una enfermedad. De este modo permite escoger la prueba óptima para su uso.
Evaluaciones económicas	En adultos mayores de 60 años, ¿cuál es la costo-efectividad de la terapia de reemplazo renal comparado con la terapia paliativa? • *Población:* adultos > 60 años • *Intervención:* terapia de reemplazo renal • *Control:* terapia paliativa • *Desenlace:* costo-efectividad, costo-utilidad, costo-beneficio, u otras	Evaluar las intervenciones teniendo en cuenta sus costos. Esto permite invertir de la mejor manera los recursos disponibles.
Prevalencia y / o incidencia	¿Cuál es la prevalencia/incidencia de desnutrición en niños menores de 5 años procedentes de la sierra del Perú? • *Población:* niños menores de 5 años con desnutrición en la sierra del Perú • *Desenlace:* prevalencia/incidencia de desnutrición	Evaluar la frecuencia de la carga de la enfermedad. Esto permite evaluar los cambios y tendencias de las enfermedades a lo largo del tiempo.
Basado en la experiencia (cualitativo)	¿Cuál es la percepción de la calidad de atención en adultos de países de altos ingresos? • *Población:* adultos • *Fenómeno de interés:* percepción de calidad de atención • *Contexto:* en países de altos ingresos	Evaluar la experiencia o el significado de un fenómeno particular enfocado desde la perspectiva los individuos.

1. Formular una pregunta de investigación.
2. Realizar la búsqueda sistemática.
3. Seleccionar los estudios que hayan respondido a la pregunta establecida.
4. Extraer los datos de interés de los estudios seleccionados.
5. Valorar el riesgo de sesgo de los estudios seleccionados.
6. Cuando sea pertinente, realizar la síntesis cuantitativa de los resultados - conocida como meta-análisis (MA).
7. Evaluar el sesgo de reporte.
8. Evaluar la certeza de la evidencia.

Para mayor información sobre estos pasos, sugerimos leer el "Manual Cochrane de revisiones sistemáticas de intervenciones" [6].

Al leer una RS, es necesario tener en cuenta dos aspectos importantes: 1) que los resultados de la RS dependerán de la calidad y características del cuerpo de la evidencia; es decir, del conjunto de los estudios primarios incluidos y 2) que no todas las RS publicadas siguieron una metodología adecuada, ya sea por inexperiencia de los autores, por apuro por publicar pronto, o incluso debido a intereses secundarios de favorecer o no cierta intervención [7]. Por ello, es importante que los profesionales de la salud no se limiten a creer en las conclusiones de una RS, sino que posean las competencias necesarias para evaluarlas críticamente.

PRESENTACIÓN DE CASO CLÍNICO

Para guiar al lector por el proceso de lectura de RS haremos uso de un ejemplo ficticio: imagina que eres un médico internista a cargo del manejo de un paciente con cierta condición que lo predispone a comer insaciablemente, y por tanto incrementa sus posibilidades de morir, a la que llamaremos "síndrome del glotón" (SDG). Este paciente te consulta sobre un tratamiento (sustancia X), que podría ayudarlo a comer menos y así reducir su índice de masa corporal (IMC) y su mortalidad.

En este caso, decides tomar una decisión basada en evidencias, para lo cual seguirás a grandes rasgos los siguientes pasos [8]:

1. Formular la pregunta.
2. Buscar la evidencia que conteste tu pregunta (en este caso, buscar una RS).
3. Evaluar e interpretar la RS.
4. Aplicar o no en el paciente la intervención sugerida.
5. Evaluar los resultados obtenidos.

La pregunta planteada es: En personas con SDG, ¿cuáles son los beneficios y los daños de brindar la sustancia X? Siguiendo el formato PICO: Población: personas con SDG, intervención: sustancia X, control: no brindar la sustancia X. Desenlace: beneficios y daños. El presente artículo se enfocará el tercer paso (evaluar e interpretar la RS).

Luego de realizar la búsqueda bibliográfica, has encontrado un estudio que parece ser una RS que responde a la pregunta de interés. Para saber si se trata de una RS, a veces es suficiente con mirar el título -donde se explicita que el estudio es un "systematic review" o "meta-analysis"-, aunque no todas lo tienen evidente. Si tienes dudas, revisas el resumen o incluso el cuerpo del artículo, donde deben mencionar que realizaron una búsqueda sistemática de artículos.

Una vez que te hayas cerciorado que estás leyendo una RS, procedes a revisar el objetivo de la RS, ya sea en el resumen o en el último párrafo de la sección "introducción", para compararlo con la pregunta que te has planteado.

LEYENDO LA SECCIÓN DE MÉTODOS

Criterios de inclusión

Para empezar a leer una RS, busca en la sección de métodos los criterios de inclusión de los estudios y evalúa si estos criterios corresponden al paciente o la situación que estás evaluando. Por ejemplo, sería problemático que tu paciente esté gestando, pero la RS incluya solo estudios realizados en pacientes no gestantes; y tú creas que la gestación podría afectar la eficacia o seguridad de la sustancia X.

Adicionalmente, en esta sección se mencionará qué diseños de estudio fueron considerados en la RS. Las RS pueden incluir ensayos clínicos aleatorizados (ECA) y/o estudios observacionales. Los ECA serían la fuente de evidencia más confiable para evaluar eficacia, por tener menor riesgo de sesgo [9,10].

Estrategias de búsqueda

A continuación, busca las fuentes de información consultadas (usualmente bases de datos como Medline, Scopus, *Cochrane Central Register of Controlled Trials* [CENTRAL], Embase, *Web of Science*, etc.) por el estudio y los términos de búsqueda usados en cada una de estas bases. Se debe evaluar si las fuentes de información y los términos son lo suficientemente sensibles para captar todos los estudios de interés.

Resto de la metodología

En el resto de la sección de métodos se explica cómo se realizó la selección de estudios y la extracción de los datos, cómo se evaluó el riesgo de sesgo de los estudios, cómo se resumieron los resultados, y otros análisis realizados como los de búsqueda de sesgo de publicación. Es importante leer todo ello para entender el proceso seguido.

LEYENDO LA SECCIÓN DE RESULTADOS

Selección de estudios

La sección de "resultados" suele comenzar presentando la selección de estudios mediante un flujograma que muestra: el número total de estudios obtenidos luego de la búsqueda sistemática, el número restante de estudios luego de eliminar

Tabla 2. Características de los estudios incluidos.

Estudio*	Tipo de ECA	País	Edad promedio - años (rango)	Población **	Intervención	Control	Seguimiento máximo	Financiamiento
Banner 2013	Paralelo	EEUU	44 (18-50)	I: 89 C: 98	Sustancia X	Placebo	1 año	No
Cage 2022	Cruzado	EEUU	34 (19-46)	I: 55 C: 57	Sustancia X	Placebo	1 año	No
Fury 2021	Paralelo	EEUU	35 (20-48)	I: 75 C: 60	Sustancia X	Placebo	1 año	Shield Group
Odison 2020	Cruzado	EEUU	45 (24-56)	I: 77 C: 84	Sustancia X	Placebo	1 año	University of Asgard
Parker 2020	Paralelo	EEUU	42 (22-48)	I: 83 C: 88	Sustancia X	Placebo	1 año	Spider Society
Stark 2022	Paralelo	EEUU	43 (18-53)	I: 880 C: 846	Sustancia X	Placebo	18 meses	No
Maximoff 2014	Paralelo	Corea	38 (24-55)	I: 64 C: 66	Sustancia X	Placebo	1 año	No
Pym 2015	Paralelo	Afganistán	39 (18-52)	I: 72 C: 74	Sustancia X	Placebo	13 meses	No
Quill 2018	Cruzado	Armenia	32 (23-47)	I: 62 C: 63	Sustancia X	Placebo	1 año	No
Rogers 2016	Cruzado	Chipre	41 (18-54)	I: 30 C: 35	Sustancia X	Placebo	1 año	Veterans Force
Romanoff 2019	Cruzado	Georgia	46 (20-50)	I: 98 C: 105	Sustancia X	Placebo	15 meses	Black Widow Research Group
Strange 2024	Paralelo	Tayikistán	46 (19-53)	I: 79 C: 88	Sustancia X	Placebo	1 año	University of Kamar-Tash
T'Challa 2025	Cruzado	Laos	48 (25-58)	I: 115 C: 120	Sustancia X	Placebo	18 meses	University of Wakanda

ECA: Ensayo clínico aleatorizado
* El desenlace para todos los estudios fue el índice de masa corporal (IMC) y la mortalidad
** I: Intervención, C: Control

los estudios duplicados, el número restante de estudios luego de evaluar su elegibilidad al revisar el título y resumen, y el número de estudios incluidos luego de evaluar su elegibilidad mediante la revisión a texto completo [11].

Para nuestro ejemplo, dicho flujograma muestra que se realizó una búsqueda de ECA en diversas bases de datos, y finalmente, se incluyó a 13 ECA (Figura 1A).

Características de los estudios incluidos

Las RS suelen presentar una tabla que resume las características de los estudios primarios incluidos. Entre la información que es relevante presentar se encuentran las características de la población, la intervención, el comparador, los desenlaces y el origen del financiamiento o los conflictos de interés [12]. Esto permite tener una idea general de qué tan comparables son los estudios - es posible que no todos evaluaron exactamente a la misma población o hayan utilizado la misma dosis de un medicamento - y si responden a nuestra pregunta y al caso particular de nuestro paciente.

Además, la información sobre el financiamiento del estudio nos permite determinar potenciales conflictos de interés: por ejemplo, es posible que los estudios financiados por empresas que vendan la sustancia X tengan algún interés en que el estudio muestre que dicha sustancia es útil, y eso podría llevar en el peor de los casos a resultados sesgados [13].

En nuestro ejemplo, se presentan las características de los estudios incluidos, incluyendo los componentes de la pregunta PICO y el financiamiento (Tabla 2).

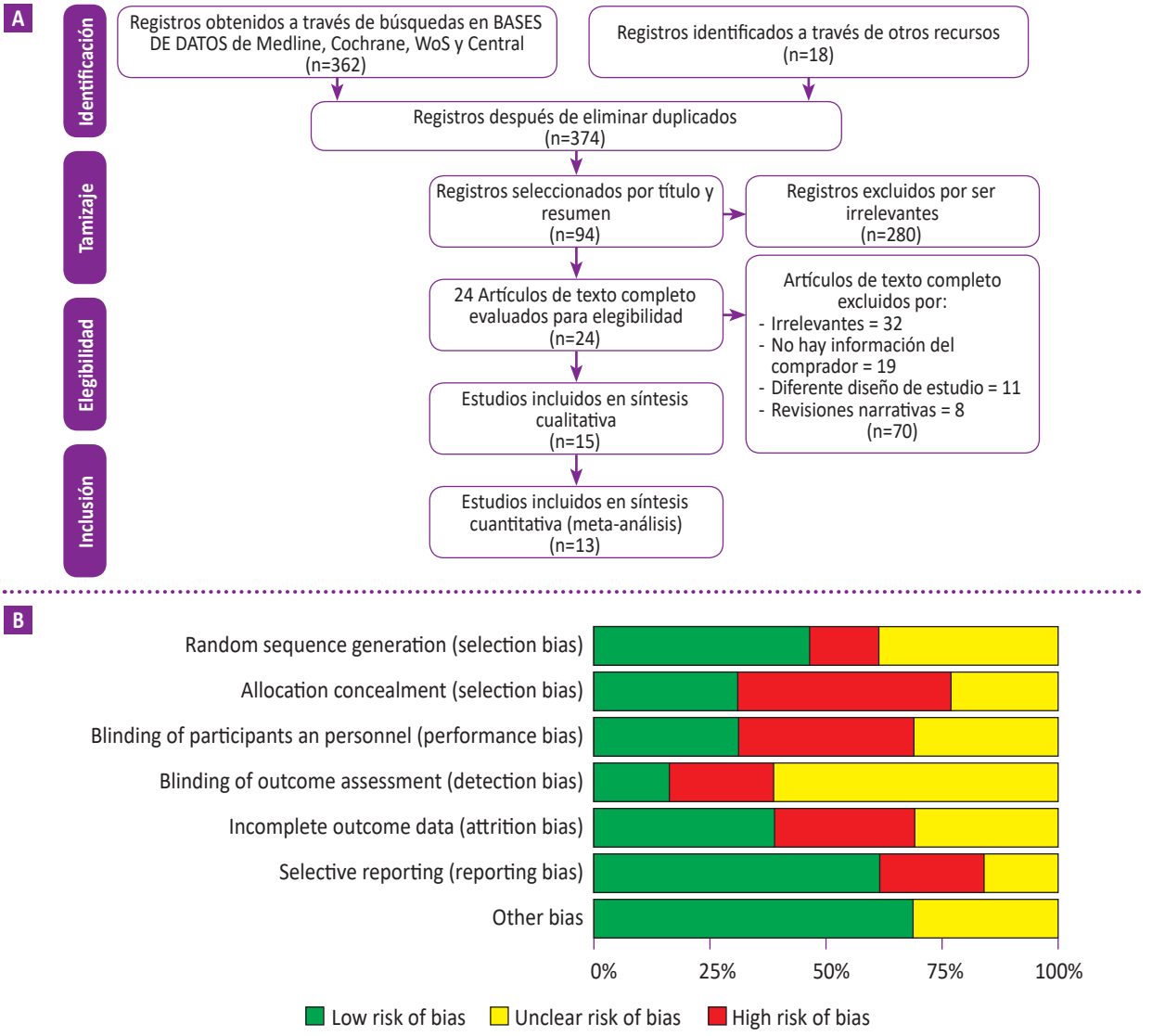

Figura 1. (A) Flujograma de selección de estudios primarios. **(B)** Evaluación de riesgo de sesgo por dominio usando la herramienta de riesgo de sesgo de Cochrane. **(C)** Evaluación de riesgo de sesgo usando la herramienta de riesgo de sesgo de Cochrane.

Riesgo de sesgo

A continuación, se presenta la evaluación del riesgo de sesgo de los estudios incluidos. Para realizar dicha evaluación, los autores de la RS utilizan alguna de las herramientas disponibles en la literatura [14]. Para evaluar este riesgo de sesgo en un ECA, la herramienta más usada es la de Cochrane de riesgo de sesgo [6], en tanto que para evaluar lo mismo en estudios observacionales se tienen varias herramientas como *Risk of bias in non-randomised studies - of interventions* (ROBINS-I) [15] o Newcastle-Ottawa [16].

La herramienta de Cochrane evalúa el riesgo de sesgo a través de seis dominios (generación de la secuencia de aleatorización, ocultación de la asignación a los grupos, cegamiento de los participantes y/o del investigador, manejo de los datos de resultados incompletos, notificación selectiva, y otros sesgos) [6]. Para cada dominio, el riesgo de sesgo puede calificarse como alto, bajo, o poco claro; categorías que son representadas por los colores rojo, verde, y ámbar, respectivamente [6,12]. Para conocer más sobre esta herramienta, recomendamos leer el capítulo 8 (*Assessing risk of bias in a randomized trial*) del "Manual Cochrane de revisiones sistemáticas de intervenciones versión 5.1.0" [6].

De manera general, si es que la gran mayoría de los círculos estarían pintados de color verde, diríamos que la evidencia tiene bajo riesgo de sesgo; mientras que si estuvieran pintados de amarillo o rojo diríamos que el riesgo es alto.

Para nuestro ejemplo, se muestran dos formas de presentar el riesgo de sesgo de los estudios (Figuras 1B y 1C). La forma más informativa es aquella que menciona el riesgo de sesgo para cada estudio (Figura 1C). Como se puede apreciar, de los trece estudios evaluados, solo los estudios de Strange y T'Challa tuvieron un bajo riesgo de sesgo en la mayoría de los dominios de la herramienta de Cochrane.

Resultados de los estudios individuales incluidos en la RS

Al llegar a esta parte, debemos preguntarnos si se están evaluando los desenlaces (ya sean beneficiosos o dañinos) más importantes para el paciente. Esto resulta importante, pues no es raro que las RS se concentren en desenlaces poco importantes, lo cual nos puede llevar a tomar una decisión sesgada.

Por ejemplo, si una RS evalúa el efecto de los fármacos reductores de fosfato en pacientes con insuficiencia renal, podríamos plantear que existen desenlaces importantes (p.e. mortalidad, fracturas, dolor), y desenlaces de menor importancia (p.e. flatulencias) para el paciente [17].

Además, existen desenlaces subrogados. Un desenlace subrogado de fracturas sería densidad ósea, pues tenemos evidencia que hay asociación entre ambos. Sin embargo, es posible que una mejora en un desenlace subrogado no produzca ningún beneficio en los desenlaces importantes para el paciente. Debemos tener todo esto en cuenta al comparar los daños y beneficios para tomar una decisión [17].

Los resultados de los estudios incluidos en la RS se pueden resumir de manera cualitativa - presentando una tabla donde se describa los resultados de cada estudio - o de manera cuantitativa mediante un meta-análisis (MA). Finalmente, de realizarse algún MA, los autores podrían evaluar el sesgo de reporte de los estudios.

En el siguiente capítulo profundizaremos en la lectura de los MA, y en el subsiguiente la evaluación de sesgo de reporte.

META-ANÁLISIS

El MA es una técnica estadística usada para resumir en un único valor los resultados de dos o más estudios que hayan comparado dos grupos (un grupo intervención y un grupo control) [4]. Para nuestro ejemplo, los estudios incluidos han comparado un grupo de pacientes que recibió la sustancia X (grupo intervención) con un grupo de pacientes que recibió un placebo (grupo control).

Para su correcta interpretación se deben entender algunos conceptos que explicaremos a continuación:

Tamaño del efecto (effect size)

El tamaño de efecto es un término genérico que indica la dirección y magnitud del efecto de una intervención, y se suele representar de manera numérica usando alguna medida de efecto de acuerdo al desenlace estudiado [4,18].

Las medidas de efecto se presentan junto con su intervalo de confianza (IC), usualmente el IC 95%. Este es un rango en el cual esperamos encontrar con una alta confianza el valor de la medida de efecto en la población. Es decir, si tenemos un riesgo relativo (RR) de 1,5 con un IC 95% de 1,2 a 1,8, entenderemos que el RR en la muestra evaluada fue de 1,5, en tanto que el RR en la población (lo que realmente nos interesa) con un 95% de confianza se ubicará entre 1,2 y 1,8 [4]. Esto se detalla en la Tabla 3.

Forest plot

El *forest plot* es un gráfico que muestra los resultados de cada estudio (estimados puntuales y sus IC) y del resumen estadístico de dichos estudios (estimado global y su IC) [18].

Para describir los componentes del *forest plot*, revisemos los gráficos de nuestro ejemplo, en los que se representan los MA realizados para los desenlaces de mortalidad (Figuras 2A y 2B) e IMC (Figuras 2C y 2D).

La primera columna (*study or subgroup*) corresponde a la identificación de cada estudio, que usualmente incluye el apellido del primer autor y el año de publicación del estudio. Las siguientes columnas (sustancia X y placebo) describen las características del grupo intervención (sustancia X) y control (placebo) en cada estudio; es decir, el número de eventos y el total de personas incluidas en cada estudio para desenlaces

Tabla 3. Formas de medir el tamaño de efecto

Desenlaces Dicotómicos Un desenlace dicotómico es aquel que solo admite dos alternativas: sí o no (por ejemplo, la mortalidad a los seis meses, pues la persona estudiada bien estará muerta o no estará muerta). Para evaluar estos desenlaces se suelen usar medidas de efecto como *risk ratio* (RR), *odds ratio* (OR), o *hazard ratio* (HR). En las medidas de efecto para desenlaces dicotómicos, por tratarse de ratios, el valor será uno "1" cuando la proporción de los individuos que presentaron el desenlace sea idéntica en el grupo intervención y el grupo control. Como consecuencia, cuando el IC de una medida de efecto para desenlaces dicotómicos incluya al valor uno, no existiría una asociación estadísticamente significativa entre el grupo (intervención/control) y el desenlace Ejemplo: se hizo un estudio para determinar la mortalidad luego de recibir un tratamiento A comparado con recibir un tratamiento B.		
Razón de riesgo o *Risk ratio* (RR)	Se calcula dividiendo el riesgo de morir del grupo que recibió el tratamiento A entre el riesgo de morir del grupo que recibió el tratamiento B.	**RR/OR/HR = 1:** El riesgo de presentar el desenlace es similar en el grupo A y B (no se encuentra asociación) **RR/OR/HR > 1:** El riesgo de presentar el desenlace es mayor en el grupo A que en el grupo B **RR/OR/HR < 1:** El riesgo de presentar el desenlace es menor en el grupo A que en el grupo B
Odds ratio (OR)	Se calcula dividiendo el odds (posibilidad) de morir del grupo que recibió el tratamiento A entre el odds (posibilidad) de morir del grupo que recibió el tratamiento B.	
Hazard Ratio (HR)	Se calcula dividiendo la tasa instantánea (hazard) de mortalidad al recibir el tratamiento A entre la tasa instantánea (hazard) de mortalidad al recibir el tratamiento B. Para este cálculo se considera el tiempo de seguimiento.	
Desenlaces Continuos Un desenlace continuo es aquel que se expresa con números (por ejemplo, la pérdida de peso, que puede evaluarse en kilogramos). Para evaluar estos desenlaces, se suele usar medidas de efecto como diferencia de medias (en inglés: *mean difference* [MD]), y diferencia de medias estandarizadas (en inglés: *standardized mean difference* [SMD]). En tanto que, en las medidas de efecto para desenlaces continuos, por tratarse de diferencias de medias, el valor será cero "0" cuando la media del desenlace sea igual en el grupo intervención y el grupo control. Como consecuencia, cuando el IC de una medida de efecto para desenlaces continuos incluya al valor cero, no existiría una asociación estadísticamente significativa entre el grupo (intervención/control) y el desenlace. Ejemplo: se hizo un estudio para determinar el dolor (usando una escala para ello) luego de recibir un tratamiento A comparado con recibir un tratamiento B.		
Diferencia de medias (DM) o *Mean difference (MD)*	Se calcula restando la media de dolor en el grupo A y menos la media del dolor en el grupo B. Si el puntaje en la escala de dolor luego de recibir el tratamiento A y B es de 20 y 14 días respectivamente, la MD es de 6 puntos.	**MD/SMD=0:** la media del desenlace en el grupo A es igual a la media del desenlace en el grupo B (no se encuentra asociación) **MD/SMD>0:** la media del desenlace en el grupo A es mayor a la media del desenlace en el grupo B **MD/SMD<0:** la media del desenlace en el grupo A es menor a la media del desenlace en el grupo B
Diferencia de medias estandarizada (DME) o *Standardized mean difference (SMD)*	Se calcula dividiendo la MD entre la desviación estándar de la variable evaluada. Se usa para poder comparar escalas con unidades de medida diferentes (por ejemplo, una escala de dolor del 1 al 100, y otra de A al E, no podrán compararse si se usa MD, pero sí usando SMD)	

dicotómico, o la media y la desviación estándar para desenlaces cuantitativos. La siguiente columna (*weight*) reporta el peso de cada estudio, que hace referencia al aporte del estudio (en porcentaje) al estimado global del MA. La siguiente columna (*risk ratio* en las Figuras 2A y 2B, o *mean difference* en las Figuras 2C y 2D) reporta numéricamente el tamaño del efecto.

Finalmente, en el extremo derecho se muestra propiamente la gráfica de *Forest Plot.* En el *forest plot*, para cada estudio primario, el estimado puntual está representado por un cuadrado cuyo tamaño es directamente proporcional al peso; es decir, aquellos estudios con mayor peso tendrán un cuadrado más grande, por ejemplo, nótese que el estudio de Stark del primer subgrupo de la Figura 2B tiene un peso mayor al resto de los estudios (53,3%) y en concordancia el tamaño del cuadrado correspondiente es notoriamente mayor. Los IC de cada estudio son representados mediante líneas horizontales cuyos extremos representan el límite inferior y superior del IC. Finalmente, el estimado global -el resultado final, que resume estadísticamente los estudios primarios - está representado por un rombo, de manera que los vértices superior e inferior representan el valor puntual, y los vértices laterales representan su IC.

En la parte inferior del *forest plot* hay una regla llamada "Escala de efecto", de la cual se traza una línea vertical que lleva por nombre "Línea de no efecto". Si el desenlace es dicotómico, la

línea de no efecto se proyectará desde el uno (Figuras 2A y 2B); si es continuo, desde el cero (Figuras 2C y 2D). Si es que una de las líneas horizontales o de los rombos cruza dicha línea, podemos decir que el efecto no es estadísticamente significativo.

Finalmente, los MA pueden evaluar no solo el resultado total, sino también subgrupos de interés. Por ejemplo, en los MA de la Figura 2, se evalúa dos subgrupos de estudios (hechos en EEUU y hechos en Asia), y se presenta un estimado global para cada subgrupo, además del estimado global total en la última fila (que incluye ambos subgrupos).

Vamos a tomar como ejemplo a la Figura 2A en la que vemos un *forest plot* para el desenlace de mortalidad. Podemos observar que se trata de un desenlace dicotómico, pues están usando RR y la línea vertical está trazada en el valor uno. Este MA tiene dos subgrupos. Considerando a los 13 estudios, vemos que en el grupo de intervención (sustancia X) 1351 de 1749 pacientes presentaron el evento (es decir, murieron); y para el grupo control (placebo) 1328 de 1784 pacientes murieron.

En esta figura, los resultados de cada subgrupo se resumen en un estimado global (un rombo) diferente, y un tercer rombo ubicado en la parte inferior resume todos los estudios en conjunto. Vamos a interpretar este último rombo, que refleja un RR de 1,06 con un IC 95% de 0,98 a 1,14 (como se explicita en la figura). El valor puntual de 1,06 se interpretaría mencionando que el grupo intervención tuvo un 6% más de muertes que el grupo control. Sin embargo, la interpretación formal debe considerar los intervalos de confianza. De esta manera, puesto que el IC incluye el valor "1" - lo cual también se puede ver en el gráfico, pues el rombo cruza la línea vertical de no efecto -, nuestra conclusión será que no encontramos diferencia estadísticamente significativa en mortalidad entre el grupo que recibió "sustancia X" y el que recibió placebo, en pacientes con SDG. Otra manera de identificar la significancia estadística del estimado global es fijarse en el test de *overall effect* situado en la parte inferior de la tabla; este también nos proporciona un valor de *p* que en nuestro ejemplo es de 0,15 (mayor que 0,05), lo cual también nos indica que el resultado no fue estadísticamente significativo.

Heterogeneidad

Si un grupo de estudios evalúan la misma pregunta PICO, se espera que los resultados sean similares entre sí, aunque es de esperar que debido al azar encontremos cierta variabilidad entre sus resultados. Sin embargo, una variabilidad muy grande puede deberse a variabilidad clínica (referida a diferencias en las características de la población, en los tipos de intervenciones, en los cuidados recibidos por el grupo de comparación, o en las mediciones de los desenlaces) y/o a variabilidad metodológica (referida a las diferencias en el diseño y en la ejecución).

La interpretación estadística de la variabilidad entre los efectos de dos o más estudios se denomina heterogeneidad [19]. La heterogeneidad entre los estudios primarios de un MA debe ser evaluada para permitir una mejor comprensión de los resultados de dicho MA. Para ello, se pueden realizar diversas estrategias, preferiblemente en conjunto [2,5,19]:

1. Observación del *forest plot.* Podremos decir que existe poca heterogeneidad cuando los estimados de los estudios se ubican cercanamente entre sí y superponen sus intervalos de confianza.
2. Interpretar medidas como el I^2 de Higgins. Para el estadístico I^2 un 0% sugiere que el azar es el responsable de la variabilidad, mientras que un 100% sugiere que la variabilidad es excesiva. Si bien no existe un punto de corte definitivo porque la importancia de la inconsistencia depende de muchos factores, Cochrane sugiere que un I^2 de hasta 40% sería lo esperado por el azar, y más allá de eso tendría otras causas [6].
3. Usar pruebas estadísticas como la prueba de chi cuadrado para heterogeneidad. La prueba chi cuadrado tiene como hipótesis nula que todos los estudios presentan el mismo efecto, de manera que según algunos Cochrane un valor de *p* menor o igual a 0,10 determinaría una diferencia significativa (una heterogeneidad) [6].
4. En caso se haya realizado un MA con modelos de efectos aleatorios, existe otra prueba conocida como Tau cuadrado (Tau^2), la cual estima la varianza entre los tamaños de efecto de los estudios en un MA. Lamentablemente, aún no existe un consenso sobre qué puntos de corte usar para interpretar esta prueba [6].

En nuestro ejemplo, para el MA mostrado en la Figura 2A podemos ver que se realizaron pruebas de heterogeneidad para cada subpoblación y el total. Revisaremos el total: el valor p de la prueba chi cuadrado de heterogeneidad es 0,001, y el I^2 es de 63%. Todo esto nos indica que existe heterogeneidad importante.

Si en caso se evidencie heterogeneidad en el MA, el manual de Cochrane [6] propone siete estrategias para abordar la heterogeneidad: 1) verifique nuevamente los datos, 2) no realice un MA, 3) explore la heterogeneidad, 4) ignore la heterogeneidad, 5) realice un MA de efectos aleatorios, 6) cambie la medida de efecto y 7) excluya estudios. La decisión final deberá ser discutida, tomada y justificada por el grupo investigador.

Modelos de efectos

Para calcular el estimado global en un MA, se puede hacer uso de diversos modelos matemáticos, los cuales se dividen en dos grupos: modelos de efectos fijos (fixed effects) y de efectos aleatorios (random effects).

Se realizará un MA usando un modelo de efectos fijos cuando se asuma que los estudios han evaluado diferentes muestras representativas de una misma población, sin cambiar el contexto ni la metodología; es decir, cuando el efecto real en la población sea el mismo en cada estudio. En estos casos, si bien se espera que el efecto de cada estudio sea diferente (por variación por principio del azar), la heterogeneidad debe ser baja ($I^2 < 40\%$).

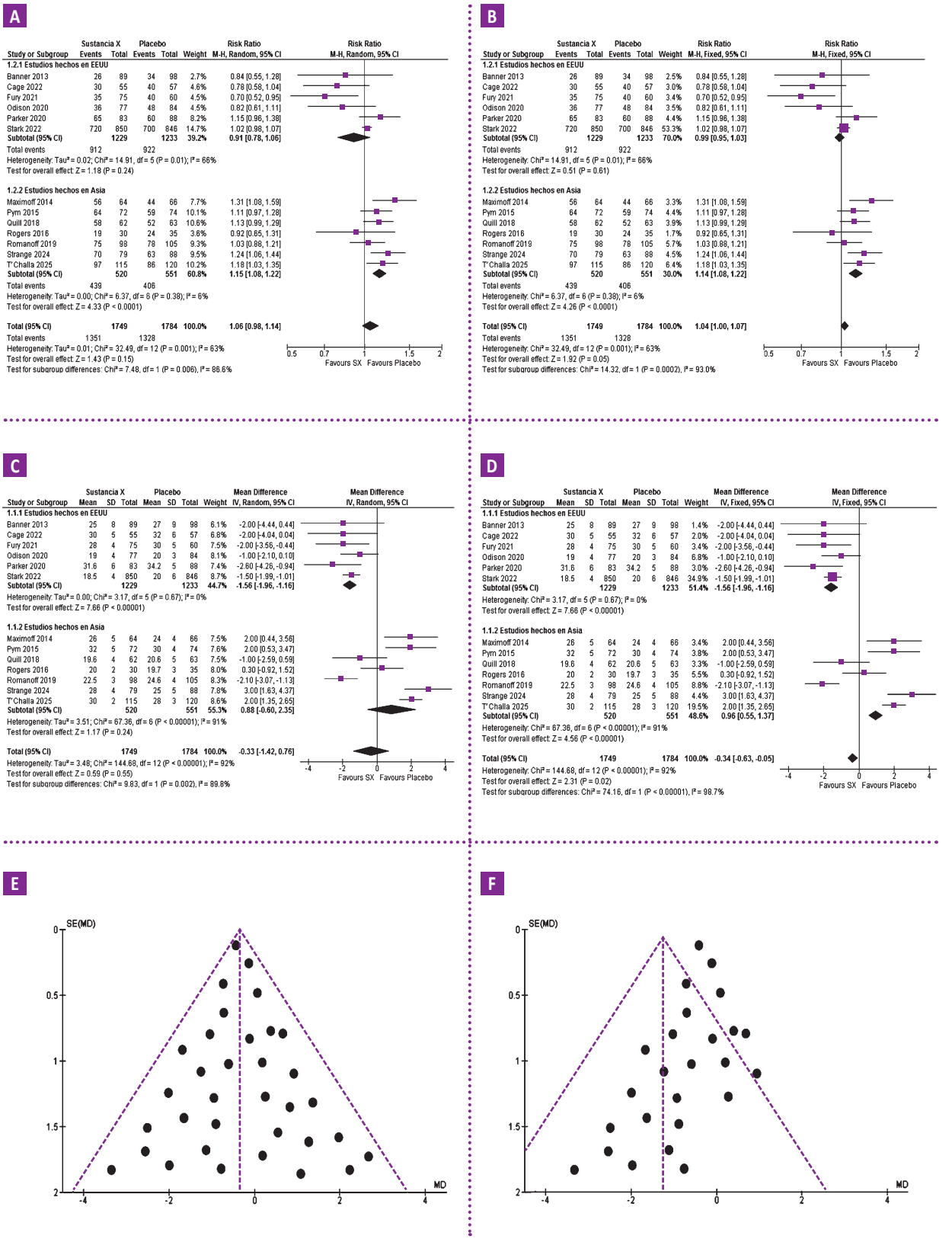

Figura 2. Comparación entre la sustancia X y Placebo. **(A)** Forest plot de mortalidad, modelo de efecto aleatorio. **(B)** Forest plot de mortalidad con modelo de efecto fijo. **(C)** Forest plot de índice de masa corporal con modelo de efecto aleatorio. **(D)** Forest plot de índice de masa corporal con modelo de efecto fijo. **(E)** Funnel plot de una variable numérica con ausencia de sesgo de publicación. **(F)** Funnel plot de mortalidad de una variable numérica con presencia de sesgo de publicación.

Se realizará un MA usando un modelo de efectos aleatorios cuando se asuma que los estudios no reflejan un solo efecto real, por poseer diferentes poblaciones, intervenciones, comparadores o formas de evaluar el desenlace. En este caso, se espera una alta heterogeneidad, por lo cual este modelo se usa cuando se obtenga un $I^2 \geq 40\%$ [20]. Los modelos aleatorios suelen obtener estimados globales con IC más amplios que los modelos fijos y mayor probabilidad de no obtener resultados estadísticamente significativos, por lo cual se dice que son más conservadores, y algunos autores prefieren usar modelos aleatorios siempre que haya alguna duda sobre el cumplimiento de los supuestos para usar modelos fijos [20].

En el ejemplo, se muestran los MA realizados con modelos de efectos aleatorios y fijos, para el desenlace de mortalidad (Figuras 2A y 2B, respectivamente) y para el desenlace de IMC (Figuras 2C y 2D, respectivamente). En todos los casos se observa heterogeneidad ($I^2 \geq 40\%$), por lo cual se optaría por un modelo de efecto aleatorio. Los modelos fijos fueron realizados solo con fines académicos.

Nótese que los estimados globales calculados con modelos de efectos aleatorios son más amplios que los calculados con efectos fijos. En algunos casos, el cambio de modelos puede variar las conclusiones, como sucedió para el desenlace de IMC, donde el MA realizado usando modelos de efectos aleatorios no fue estadísticamente significativo, en tanto que el realizado usando modelos de efectos fijos sí lo fue (Figuras 2C y 2D, respectivamente).

Análisis de subgrupo

En ciertas ocasiones queremos comparar los estimados globales en dos subgrupos de estudios [19]. Por ejemplo, en la Figura 2A, se plantearon dos subgrupos: el primero corresponde a los estudios hechos en EEUU y el segundo a los estudios hechos en Asia. Es posible que la intervención tenga un efecto diferente entre estos grupos, por diferencias genéticas, sociales, del sistema de salud, u otras.

Para comparar los resultados; es decir, para evaluar si hay diferencia de los estimados globales de los subgrupos, se calcula un *Test for subgroup differences* (que se presenta en la última fila de cada gráfico), obteniendo un valor de *p*. Cuando *p* sea menor a 0,10, se suele interpretar que existen diferencias estadísticamente significativas entre los estimados globales de los subgrupos [21].

En el ejemplo, el *Test for subgroup differences* tuvo un valor de *p* de 0,006, lo cual indica que el efecto de la intervención en EEUU fue estadísticamente diferente al efecto en Asia. En concordancia, en el *Forest Plot* se observa que, en el grupo de estudios hechos en EEUU la intervención no parece tener un efecto en el desenlace, en tanto que en el subgrupo de estudios hechos en Asia la mortalidad parece ser mayor en el grupo intervenido. En este caso, incluso se podría tomar una decisión diferente dependiendo si el paciente es de EEUU o de Asia.

Análisis de sensibilidad

Luego de realizar un MA, es posible que nos demos cuenta que algunos de los estudios incluidos son diferentes al resto, por tener mayor riesgo de sesgo, haber sido realizados en otra población, etcétera. Si es que creemos que estos estudios podrían estar modificando artificialmente el estimado global, podemos hacer un nuevo MA excluyéndolos. Este tipo de análisis se conocen como "análisis de sensibilidad", y deben ser tomados con cautela debido a que usualmente son análisis no planeados, y los autores de los MA pueden forzarlos para obtener un resultado conveniente [5,19].

VALORACIÓN DE SESGO DE REPORTE Y SESGO DE PUBLICACIÓN

El sesgo de reporte se refiere a la exclusión sistemática de estudios que responden a la pregunta PICO. Puede presentarse por sesgo de publicación - no se han publicado todos los estudios que se han realizado -, idioma, citación, entre otros [22]. Resulta de importancia, pues en varios casos se publican más los resultados significativos que los no significativos, lo cual causa que el estimado global del MA, al tomar en cuenta solo los estudios publicados con resultados significativos, sobreestime el efecto real de la intervención.

Existen diversas maneras de evaluar el sesgo de reporte, incluyendo 1) el gráfico de embudo o *funnel plot*, 2) las pruebas de Egger y de Begg, 3) el "*trim and fill*" *method*, y 4) los modelos de selección [22,23]. De ellos, los más comunes son los dos primeros. El test de Egger y el test de Begg ofrecen un valor de *p* que será interpretado como sospecha de sesgo de publicación cuando sea menor que 0,10. Entre ellos, el test de Egger es el más sensible [24,25].

El *funnel plot* es un gráfico de dispersión de los estudios que se construye a partir de un MA (Figuras 2E y 2F). Los estudios están representados por un punto. En el eje de las abscisas se presenta el tamaño del efecto de los estudios (RR, OR, HR, MD, SMD, entre otros); además, se puede apreciar que de esta se traza una línea vertical punteada que representa el valor puntual del estimado global del MA (no confundir con la línea de no efecto). De la parte superior de esta línea vertical, se proyectan dos líneas oblicuas hacia la derecha e izquierda, las cuales solo sirven de referencia para formar un embudo invertido (*funnel* en inglés). En el eje de las ordenadas se presenta el error estándar, de manera que los estudios con menor error estándar estén en la parte superior del gráfico. Cabe recordar que el error estándar está inversamente relacionado con el tamaño de la población evaluada [6,26,27].

Debido al error aleatorio, los resultados de los estudios individuales deberían estar uniformemente distribuidos alrededor de la línea vertical que muestra el efecto global, tomando idealmente una distribución en forma de embudo (*funnel*). Si es que esta distribución no es simétrica, se puede sospechar de un sesgo de reporte. Para que la evaluación del

funnel plot sea confiable, se recomienda que en el análisis se incluyan por lo menos 10 estudios [22].

Supongamos otra RS ficticia en la que se evaluará el sesgo de reporte para un desenlace numérico. En las Figuras 2E y 2F, en el eje de las abscisas se expresa la MD y en el eje de las ordenadas se expresa el error estándar. Si no existe sesgo de reporte, se espera que los estudios se distribuyan de manera simétrica alrededor del estimado global, tal como se presenta en la Figura 2E. Sin embargo, supongamos que algunos estudios pequeños que no favorecían la hipótesis del efecto protector de la intervención (los que se encuentran en la parte inferior derecha del *funnel plot*) finalmente no fueron publicados, con lo cual se obtendría el *funnel plot* de la Figura 2F, que muestra una evidente asimetría sugerente de sesgo de reporte [22,28]. Nótese que el sesgo de reporte puede alterar el estimado global, como se evidencia en la Figura 2F en la cual el valor puntual del estimado global se aleja del valor de no efecto (cero) en comparación con el *funnel plot* que incluye a todos los estudios (Figura 2E).

TABLAS DE RESUMEN DE EVIDENCIAS

Las RS actuales están presentando tablas de resumen de evidencia o *Summary of Findings* (SoF), que presentan los estimados globales su certeza (el grado de confianza en el estimado global) para cada desenlace.

La evaluación de la certeza se realiza utilizando la metodología *Grading of Recommendations Assessment, Development, and Evaluation* (GRADE) [22]. Para cada desenlace, se establece alguna de las cuatro posibles categorías de certeza: alta, moderada, baja, o muy baja. Las definiciones de cada una se encuentran casi al final de la Tabla 4.

Para evaluar la certeza de la evidencia, se considera que los MA de ECA inician con nivel de certeza alta y los MA de estudios observacionales inician con nivel de certeza baja. Luego, se evalúan cinco criterios para disminuir el nivel de certeza inicial, 1) riesgo de sesgo [23], 2) heterogeneidad [28], 3) evidencia indirecta (si la pregunta PICO del MA es similar al objetivo de la RS) [29], 4) imprecisión (si los intervalos de confianza del estimado global son amplios) [19], y 5) sesgo de publicación [19]. En MA de estudios observacionales, también se evalúan tres criterios que permiten aumentar el nivel de certeza inicial, 1) tamaño del efecto, 2) dosis-respuesta, y 3) el efecto de la posible confusión residual [30].

En nuestro ejemplo, la Tabla 4 (SoF) describe la evidencia empezando por el título de la comparación, seguido de la población, intervención y comparador de la pregunta PICO. A continuación, se muestran dos filas correspondientes a dos desenlaces (IMC y mortalidad). Cada fila corresponde a un MA.

En la primera columna se muestra el nombre de cada desenlace, le continúan los efectos absolutos anticipados con un IC 95%, el riesgo relativo (que solo pudo ser calculado por el desenlace dicotómico de mortalidad), el número de estudios meta-analizados y el número total de participantes en dichos estudios, la certeza de la evidencia, y algunos comentarios acerca del desenlace. La explicación del cálculo de la certeza se aprecia al pie de la tabla.

Al leer el resumen de la tabla SoF para IMC, podremos decir que luego de un año de seguimiento el grupo que recibió sustancia X tuvo un IMC 0,33 kg/m^2 menor al grupo que recibió placebo, pero esta diferencia no fue estadísticamente significativa (por tener un IC entre -1,42 a +0,76), resultado que tuvo una muy baja certeza de la evidencia.

Al leer el resumen de la tabla SoF para mortalidad, podremos decir que luego de un año de seguimiento el grupo que recibió sustancia X tuvo un riesgo de morir 6% mayor al grupo que recibió placebo, aunque esto no fue estadísticamente significativo, resultado que tuvo una muy baja certeza de la evidencia.

¿PUEDO CONFIAR EN ESTA RS?

Es posible que hayamos encontrado más de una RS que conteste nuestra pregunta. Para elegir cuál deberíamos usar para tomar decisiones - o si no deberíamos usar ninguna -, debemos evaluar la antigüedad y la calidad metodológica de las RS que encontremos. Por ello, se desarrollaron herramientas que evalúan críticamente la metodología de realización de una RS como *A MeaSurement Tool to Assess systematic Reviews* (AMSTAR-II) [31] o la herramienta de lectura crítica de *Critical Appraisal Skills Programme español* (CASPe) [32]. No se debe confundir estas herramientas con el checklist *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA), el cual sirve de guía para la redacción de RS, pero no evalúa su rigurosidad metodológica [33].

CONCLUSIÓN

En el presente artículo hemos repasado los puntos básicos para la lectura de RS bajo el enfoque de toma de decisiones. Esto incluyó cómo leer la sección de métodos (con énfasis en criterios de inclusión, estrategias de búsqueda) y la de resultados (selección de estudios, características de los estudios, riesgo de sesgo, meta-análisis, y valoración del sesgo de reporte), así como la interpretación de las tablas de resumen de evidencias.

Agradecimientos: agradecemos a Gandy Dolores-Maldonado por la revisión crítica realizada al presente estudio.

Contribución de los autores: Todos los autores participaron en la concepción del artículo y en su redacción. Todos los autores aprobaron la versión final del artículo.

Fuente de financiamiento: el presente artículo ha sido autofinanciado por los autores.

Conflictos de interés: los autores declaran no tener conflictos de interés con respecto al presente artículo.

Tabla 4. Tabla SOF del ejemplo.

Sustancia X comparado con Placebo para el Síndrome del glotón						
Paciente o población: Pacientes con el Síndrome del glotón Configuración: Hospital Intervención: Sustancia X Comparación: Placebo						
Desenlaces (outcomes)	**Efectos absolutos anticipados * (95% CI)**		**Efecto relativo (95% CI)**	**№ de participantes (Estudios)**	**Certeza de la evidencia (GRADE)**	**Comentarios**
	Placebo	**Sustancia X**				
Índice de masa corporal (IMC) seguimiento: media 1 años	La media índice de masa corporal era 25,6 kg/m^2	La media índice de masa corporal en el grupo de intervención fue 0,33 kg/m^2 menor (1,42 menor a 0,76 mayor)	-	3 533 (13 ECA)	⊕◯◯◯ MUY BAJA [a,b,c,d]	En el subgrupo de los estudios hechos en EEUU se encontró diferencia estadísticamente significativa, sin embargo en el análisis global no se encontró diferencia estadísticamente significativa.
Mortalidad seguimiento: media 1 años	744 por 1.000	789 por 1.000 (730 a 849)	RR 1,06 (0,98 a 1,14)	3 533 (13 ECA)	⊕◯◯◯ MUY BAJA [a,b,d,e]	En el subgrupo de los estudios hechos en Asia se encontró diferencia estadísticamente significativa. Sin embargo en el análisis global no se encontró diferencia estadísticamente significativa.

* El riesgo en el grupo de intervención (y su intervalo de confianza del 95% se basa en el riesgo asumido en el grupo de comparación y en el efecto relativo de la intervención (y su intervalo de confianza del 95%).
CI: intervalo de confianza ; MD: diferencia media; RR: razón de riesgo

Certeza de la evidencia de acuerdo al Grupo de trabajo GRADE:
Alta certeza: Estamos muy seguros de que el efecto real se aproxima al de la estimación del efecto.
Moderada certeza: Confiamos moderadamente en la estimación del efecto: es probable que el efecto real esté cerca de la estimación del efecto, pero existe la posibilidad de que sea sustancialmente diferente.
Baja certeza: Nuestra confianza en la estimación del efecto es limitada: el efecto real puede ser sustancialmente diferente de la estimación del efecto.
Muy baja certeza: Tenemos muy poca confianza en la estimación del efecto: es probable que el efecto real sea sustancialmente diferente de la estimación del efecto.

Explicaciones:
a. Nivel de riesgo de sesgo crítico
b. Heterogeneidad con valor de I^2 mayor del 40%
c. El IC95% fue muy impreciso
d. Asimetría en la gráfica de Funnel Plot
e. El IC95% incluye el valor de 1,25

REFERENCIAS BIBLIOGRÁFICAS

1. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Med Res Methodol. 2018;18(1):5.
2. Murad MH, Montori VM, Ioannidis JP, 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.
3. Gisberta J, Bonfill X. ¿Cómo realizar, evaluar y utilizar revisiones sistemáticas y metaanálisis? Gastroenterol Hepatol. 2004;27(3):129-49.
4. Israel H, Richter RR. A guide to understanding meta-analysis. J Orthop Sports Phys Ther. 2011;41(7):496-504.
5. Uman L. Systematic reviews and meta-analyses. J Can Acad Child Adolesc Psychiatry. 2011;20(1):57.
6. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Disponible en: www.handbook.cochrane.org
7. Mallett R, Hagen-Zanker J, Slater R, Duvendack M. The benefits and challenges of using systematic reviews in international development research. J Dev Effect. 2012;4(3):445-55.

8. Gordon G, Drummond R, Maureen OM, Deborah JC. Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice. Third Edition. McGraw-Hill Education; 2014.

9. Faraoni D, Schaefer S. Randomized controlled trials vs. observational studies: why not just live together? BMC Anesthesiol. 2016;16(1):102.

10. Peinemann F, Tushabe DA, Kleijnen J. Using Multiple Types of Studies in Systematic Reviews of Health Care Interventions – A Systematic Review. Plos One. 2013;8(12):e85035.

11. Urrútia G. Declaración PRISMA una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. Med Clin (Barc). 2010;135(11):507-11.

12. Centro Cochrane Iberoamericano (trad.). Manual Cochrane de revisiones sistemáticas de intervenciones, versión 5.1.0 (actualizada en marzo de 2011). Barcelona, Centro Cochrane Iberoamericano; 2012.

13. Mandrioli D, Kearns C, Bero L. Relationship between Research Outcomes and Risk of Bias, Study Sponsorship, and Author Financial Conflicts of Interest in Reviews of the Effects of Artificially Sweetened Beverages on Weight Outcomes: A Systematic Review of Reviews. Plos One. 2016;11(9):e0162198.

14. Page MJ, McKenzie JE, Higgins JPT. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. BMJ Open. 2018;8(3):e019703.

15. Sterne J, Hernán M, Reeves B, Savović J, Berkman N, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.

16. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses [Internet]. Ottawa: Ottawa Hospital Research Institute; 2013 [citado el 12 de octubre de 2018]. Disponible en: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

17. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol. 2011;64(4):395-400.

18. John Wiley and Sons. Glossary of terms in the Cochrane Collaboration, version 4.2.5.May 2005.

19. Guyatt G, Oxman A, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol. 2011;64(12):1283-93.

20. Serghiou S, Goodman SN. Random-effects meta-analysis: Summarizing evidence with caveats. JAMA. 2019;321(3):301-2.

21. Richardsona M, Garnera P, Donegan S. Interpretation of subgroup analyses in systematic reviews: A tutorial. Clin Epidemiol Glob Health. 2019;7(2):192-8..

22. Guyatt G, Oxman A, Akl E, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383-94.

23. Guyatt G, Oxman A, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol. 2011;64(4):407-15.

24. Shi X, Nie C, Shi S, Wang T, Yang H, Zhou Y, et al. Effect Comparison between Egger's Test and Begg's Test in Publication Bias Diagnosis in Meta-Analyses: Evidence from a Pilot Survey. Int J Res Stud Biosci. 2017;5(5):14-20.

25. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629.

26. Sedgwick P. Meta-analyses: how to read a funnel plot. BMJ. 2013;346:f1342.

27. Sedgwick P, Marston L. How to read a funnel plot in a meta-analysis. BMJ. 2015;351:h4718.

28. Guyatt G, Oxman A, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J Clin Epidemiol. 2011;64(12):1294-302.

29. Guyatt G, Oxman A, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol. 2011;64(12):1303-10.

30. Guyatt G, Oxman A, Sultan S, Glasziou P, Akl E, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol. 2011;64(12):1311-6.

31. Shea B, Reeves B, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008.

32. Critical Appraisal Skills Programme Español. Instrumentos para la lectura crítica [Internet]. Alicante, España: CASPe; 2016 [citado el 12 de octubre de 2018]. Disponible en: http://www.redcaspe.org/herramientas/instrumentos

33. Moher D, Liberati A, Tetzlaff J, Altman DG, The PG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Plos Medicine. 2009;6(7):e1000097.