

Chapter 16 / Capítulo 16

Critical reading of scientific articles on
health: A practical guide (Spanish Edition)
ISBN: 978-9915-9928-3-9
DOI: 10.62486/978-9915-9928-3-9.Ch16
Pages: 263-298



© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

Systematic Reviews and Meta-Analyses

Revisiones Sistemáticas y Metaanálisis

José Bernardo Antezana-Muñoz  

Adolfo Israel Vasquez Cuellar  

Freddy Ednildon Bautista-Vanegas  

Eloy Paycho Anagua  

1 Centro de Alta Complejidad Cardiovascular Presidente Juan Domingo Perón. Argentina

2 Hospital Materno Infantil, Caja Nacional de Salud (CNS). La Paz, Bolivia.

3 Neurokliniken Beelitz GmbH Neurologische Rehabilitationsklinik: Beelitz Heilstätten, Brandenburg, DE. Germany.

4 Neurologisches Fachkrankenhaus für Bewegungsstörungen/Parkinson: Beelitz Heilstätten, Brandenburg, DE. Germany.

5 Hospital de Tercer Nivel Dr. Hernán Messuti Ribera. Pando, Bolivia.

ABSTRACT

Introduction: Systematic reviews and meta-analyses represent the highest level of evidence in Evidence-Based Medicine, rigorously, transparently, and reproducibly integrating the results of multiple primary studies to answer specific clinical questions. Their application reduces selection bias, increases the precision of estimates, and strengthens clinical and healthcare decision-making. **Development:** This chapter describes the essential methodological steps for the development and critical interpretation of systematic reviews and meta-analyses, including formulating the PICO question, registering the protocol, conducting a comprehensive literature search, selecting and extracting data, assessing the risk of bias, performing qualitative and quantitative synthesis, analyzing heterogeneity, addressing publication bias, and evaluating the certainty of evidence using GRADE. It also presents tools for interpreting forest plots, funnel plots, and network meta-analyses, highlighting their usefulness in generating robust and clinically applicable evidence. **Conclusions:** Systematic reviews and meta-analyses are indispensable tools for transforming scientific evidence into well-founded clinical decisions. However, its true value depends on methodological rigor, critical appraisal of study quality, and the proper interpretation of the magnitude and certainty of the effects. A thoughtful reading allows us to distinguish between statistical significance and clinical relevance, promoting safer, more efficient, and patient-centered healthcare practices.

Keywords: Systematic Review; Meta-analysis; Evidence-Based Medicine; Publication Bias; GRADE System

RESUMEN

Introducción: Las revisiones sistemáticas y los metaanálisis representan el nivel más alto de evidencia en la Medicina Basada en la Evidencia, al integrar de forma rigurosa, transparente y reproducible los resultados de múltiples estudios primarios para responder preguntas clínicas específicas. Su aplicación reduce el sesgo de selección, incrementa la precisión de las estimaciones y fortalece la toma de decisiones clínicas y sanitarias. **Desarrollo:** El capítulo describe las etapas metodológicas esenciales para la elaboración e interpretación crítica de revisiones sistemáticas y metaanálisis, incluyendo la formulación de la pregunta PICO, el registro del protocolo, la búsqueda bibliográfica exhaustiva, la selección y extracción de datos, la evaluación del riesgo de sesgo, la síntesis cualitativa y cuantitativa, el análisis de heterogeneidad, el sesgo de publicación y la valoración de la certeza de la evidencia mediante GRADE. De esta manera, expone herramientas para interpretar forest plots, funnel plots y metaanálisis en red, destacando su utilidad para generar evidencia robusta y clínicamente aplicable. **Conclusiones:** Las revisiones sistemáticas y los metaanálisis constituyen herramientas indispensables para transformar la evidencia científica en decisiones clínicas fundamentadas. No obstante, su verdadero valor depende de la rigurosidad metodológica, la evaluación crítica de la calidad de los estudios y la adecuada interpretación de la magnitud y certeza de los efectos. Una lectura reflexiva permite distinguir entre significancia estadística y relevancia clínica, favoreciendo una práctica asistencial más segura, eficiente y centrada en el paciente.

Palabras clave: Revisión Sistemática; Metaanálisis; Medicina Basada en la Evidencia; Sesgo de Publicación; Sistema GRADE

1. CASO CLÍNICO

Un médico especialista en Medicina Interna atiende a un paciente de 55 años con diagnóstico de diabetes mellitus tipo 2 y obesidad. Durante la consulta, el paciente refiere haber consultado diversas fuentes de información científica y de divulgación médica, en las cuales se señala que los inhibidores del cotransportador sodio-glucosa tipo 2 (iSGLT2) no solo mejoran el control glucémico, sino que también reducen la mortalidad cardiovascular, las hospitalizaciones por insuficiencia cardíaca y otros desenlaces cardiovasculares mayores. Ante esta información, plantea una pregunta clínicamente relevante: ¿la evidencia disponible demuestra que estos beneficios son superiores a los obtenidos con otros antidiabéticos orales o con placebo?

El médico reconoce que, durante la última década, múltiples ensayos clínicos aleatorizados de gran tamaño entre ellos EMPA-REG OUTCOME, CANVAS Program, DECLARE-TIMI 58 y VERTIS-CV han evaluado la eficacia y seguridad cardiovascular de los iSGLT2 en diferentes poblaciones de pacientes con diabetes mellitus tipo 2. Sin embargo, también advierte que, aunque los estudios muestran una tendencia consistente hacia el beneficio, existen diferencias en la magnitud del efecto, la significancia estadística de algunos desenlaces, las características basales de las poblaciones incluidas y los criterios metodológicos empleados. Estas discrepancias dificultan establecer una conclusión definitiva a partir de un único ensayo clínico.

Frente a este escenario, surgen varias interrogantes propias de la práctica de la Medicina Basada en la Evidencia:

- ¿Cómo puede integrarse de manera rigurosa la totalidad de la evidencia disponible para responder esta pregunta clínica?
- ¿Qué importancia relativa debe otorgarse a cada ensayo clínico considerando sus diferencias metodológicas, tamaño muestral y riesgo de sesgo?
- ¿Es posible obtener una estimación global del efecto terapéutico que incremente la precisión de los resultados y reduzca la incertidumbre inherente a los estudios individuales?
- ¿Cuál es el nivel de certeza de la evidencia que sustenta la recomendación clínica?

Para responder estas interrogantes, el médico decide recurrir a una revisión sistemática con metaanálisis, considerada el diseño de investigación secundaria con mayor nivel de evidencia para sintetizar los resultados de múltiples estudios primarios. Mediante una metodología estructurada, transparente y reproducible, este enfoque permite identificar, seleccionar, evaluar críticamente y combinar cuantitativamente la evidencia disponible, disminuyendo el riesgo de sesgos, aumentando la precisión de las estimaciones del efecto y proporcionando una base más sólida para la toma de decisiones clínicas.

Como primer paso, formula la pregunta de investigación utilizando el modelo PICO, ampliamente recomendado para estructurar preguntas clínicas de intervención:

- P (Población): adultos con diabetes mellitus tipo 2.
- I (Intervención): tratamiento con inhibidores del cotransportador sodio-glucosa tipo 2 (iSGLT2).
- C (Comparación): placebo u otros fármacos antidiabéticos.
- (Outcomes o desenlaces): mortalidad cardiovascular, hospitalización por insuficiencia cardíaca y eventos cardiovasculares mayores (Major Adverse Cardiovascular Events, MACE).

Con base en esta formulación, el internista plantea la hipótesis de que una revisión sistemática acompañada de un metaanálisis de ensayos clínicos aleatorizados permitirá obtener una estimación global más precisa del efecto de los iSGLT2 sobre los principales desenlaces cardiovasculares, establecer la consistencia de la evidencia disponible y valorar el grado de certeza de los resultados, proporcionando así un fundamento científico robusto para orientar la toma de decisiones en la práctica clínica basada en la evidencia.

Tabla 1. Comparación entre ensayo clínico individual, revisión sistemática y metaanálisis

Característica	Ensayo clínico individual (ECA)	Revisión sistemática	Metaanálisis
Tipo de estudio	Investigación primaria, prospectiva, que evalúa la eficacia o seguridad de una intervención en un grupo específico de pacientes.	Investigación secundaria, que recopila, evalúa y sintetiza de forma estructurada toda la evidencia disponible sobre una pregunta clínica.	Técnica estadística que forma parte de una revisión sistemática y combina cuantitativamente los resultados de múltiples estudios.
Objetivo principal	Determinar si una intervención es eficaz y segura en comparación con un control (placebo u otro tratamiento).	Ofrecer una visión integral y crítica de la evidencia, reduciendo sesgos mediante criterios predefinidos.	Incrementar la precisión del estimador de efecto y explorar la consistencia entre estudios.
Unidad de análisis	Pacientes o sujetos incluidos en el estudio.	Estudios publicados y no publicados que cumplen con criterios de inclusión.	Resultados cuantitativos de los estudios individuales (ej. riesgo relativo, odds ratio, hazard ratio).

Característica	Ensayo clínico individual (ECA)	Revisión sistemática	Metaanálisis
Nivel de evidencia	Alto, pero limitado a la población y condiciones del estudio.	Superior, al integrar múltiples estudios y evaluar su calidad metodológica.	Muy alto, siempre que los estudios incluidos sean de buena calidad y suficientemente homogéneos.
Fortalezas	Control riguroso de variables, aleatorización y seguimiento prospectivo.	Minimiza el sesgo de selección y ofrece síntesis estructurada de la literatura.	Aumenta la potencia estadística, permite detectar efectos pequeños y realizar análisis de subgrupos.
Limitaciones	Puede no ser generalizable; resultados aislados y a veces contradictorios.	Depende de la calidad de los estudios incluidos y del rigor metodológico.	Sensible a la heterogeneidad y al sesgo de publicación; no corrige debilidades de los estudios primarios.
Aplicación clínica	Informa decisiones en contextos específicos, pero no siempre extrapolables.	Informa guías clínicas, políticas de salud y decisiones basadas en el conjunto de la evidencia.	Permite cuantificar el beneficio o riesgo promedio y explorar la consistencia de los hallazgos.

Fuente: Elaboración propia.

2. OBJETIVOS DE UNA REVISIÓN SISTEMÁTICA Y METAANÁLISIS

Los objetivos principales pueden resumirse en los siguientes ejes:

Sintetizar de manera estructurada y transparente la evidencia disponible: Una revisión sistemática recoge, evalúa y organiza la información publicada (y no publicada) sobre una pregunta clínica específica. El objetivo es reducir el sesgo, evitando que la selección de estudios dependa de la preferencia del investigador.

Cuantificar el efecto de una intervención o exposición: El metaanálisis combina estadísticamente los resultados de los estudios incluidos. Esto permite obtener una estimación puntual y su intervalo de confianza, que representa de manera más precisa el efecto global que cualquiera de los estudios individuales.

Incrementar la potencia estadística y la precisión de los hallazgos: Al unir los datos de múltiples ensayos, se aumenta la capacidad de detectar efectos pequeños, pero clínicamente relevantes. Esto es especialmente útil cuando los estudios individuales tienen muestras pequeñas o resultados aparentemente contradictorios. **Explorar la heterogeneidad entre estudios:** Un metaanálisis no solo ofrece un “resultado promedio”, sino que permite analizar si los efectos difieren entre subgrupos (ej., hombres vs. mujeres, pacientes jóvenes vs. mayores, poblaciones con distintas comorbilidades). De esta manera, ayuda a comprender qué intervenciones funcionan mejor en qué contextos.

Evaluar la calidad y certeza de la evidencia: Una revisión sistemática rigurosa incorpora herramientas como GRADE o AMSTAR 2, que valoran la calidad de los estudios y la confianza en los resultados. El objetivo es que las conclusiones no solo sean claras, sino también fiables y aplicables a la práctica clínica. **Identificar vacíos de conocimiento e impulsar nuevas investigaciones:** Al mapear lo que existe y lo que falta, las revisiones sistemáticas orientan a los investigadores hacia preguntas aún no resueltas o hacia áreas donde los resultados siguen siendo inciertos.

3. INTRODUCCIÓN

En la práctica clínica diaria, los profesionales de la salud se enfrentan constantemente a preguntas que requieren respuestas fundamentadas en la mejor evidencia disponible. Sin embargo, la literatura científica crece de manera exponencial: cada año se publican miles de ensayos clínicos, estudios observacionales y reportes en distintas áreas de la medicina. Este volumen de información hace que sea difícil, e incluso imposible, para un médico individual leer y analizar críticamente todos los estudios relevantes sobre un tema específico.

En este contexto, las revisiones sistemáticas y los metaanálisis se han convertido en herramientas fundamentales para la medicina basada en la evidencia (MBE). Una revisión sistemática permite reunir, evaluar y sintetizar, de forma estructurada y reproducible, todos los estudios pertinentes a una pregunta clínica claramente definida. De esta manera, se busca minimizar el sesgo y proporcionar una visión integral de la evidencia.

El metaanálisis, por su parte, aplica técnicas estadísticas que permiten combinar cuantitativamente los resultados de los estudios individuales. Esto no solo aumenta la precisión de las estimaciones del efecto de una intervención o exposición, sino que también mejora la capacidad de detectar diferencias sutiles que podrían pasar inadvertidas en estudios aislados.

La importancia de estas metodologías trasciende la investigación académica: sus conclusiones constituyen la base de las guías de práctica clínica, orientan políticas de salud pública y permiten a médicos, pacientes y gestores tomar decisiones informadas. Al mismo tiempo, identifican áreas donde la evidencia es débil o insuficiente, señalando así la necesidad de nuevas investigaciones.

A lo largo de este capítulo se explorarán en detalle los fundamentos, pasos metodológicos, herramientas estadísticas, fuentes de sesgo y aplicaciones clínicas de las revisiones sistemáticas y metaanálisis. El objetivo es que el lector no solo comprenda cómo interpretar sus resultados, sino que también adquiera la capacidad crítica de valorar su calidad y aplicabilidad en la práctica diaria.

4. METODOLOGÍA

La metodología de una revisión sistemática con o sin metaanálisis sigue un proceso riguroso, diseñado para reducir el sesgo y garantizar que las conclusiones sean reproducibles y útiles en la práctica clínica. Aunque existen variantes según el tipo de revisión, los pasos fundamentales pueden resumirse en los siguientes:

4.1 DEFINICIÓN DE LA PREGUNTA DE INVESTIGACIÓN

Al analizar críticamente una revisión sistemática, el primer aspecto que debe evaluarse es si la pregunta de investigación está claramente definida y estructurada. Este punto inicial orienta toda la revisión: qué estudios se buscan, cuáles se incluyen y cómo se interpretan los hallazgos.

Las revisiones de calidad suelen utilizar el modelo PICO (Population, Intervention, Comparator, Outcome), ampliamente recomendado por el Cochrane Handbook y las guías PRISMA 2020. Una pregunta PICO bien planteada asegura que la revisión sea específica, reproducible y clínicamente relevante.

Aspectos que debe valorar el lector crítico:

- Claridad: ¿La población, la intervención, el comparador y los desenlaces están bien delimitados?
- Relevancia clínica: ¿Los desenlaces son realmente importantes para la práctica (ej. mortalidad, calidad de vida) o se limitan a variables intermedias?
- Pertinencia: ¿La pregunta responde a una necesidad clínica concreta o es demasiado amplia/difusa?
- Adecuación del diseño: ¿El modelo PICO u otra variante (PICOTS, PEO) se ajusta al tipo de revisión (tratamiento, diagnóstico, pronóstico, cualitativa)?

Para nuestro ejemplo: Un paciente con diabetes mellitus tipo 2 consulta a su médico sobre si los inhibidores de SGLT2 realmente reducen la mortalidad cardiovascular en comparación con otros antidiabéticos. Al buscar evidencia, el médico encuentra una revisión sistemática con metaanálisis que formula su pregunta así:

- P (Población): Adultos con diabetes tipo 2.
- I (Intervención): Tratamiento con inhibidores de SGLT2.
- C (Comparador): Placebo o antidiabéticos orales distintos a iSGLT2.
- (Outcomes): Mortalidad cardiovascular, hospitalización por insuficiencia cardíaca y eventos cardiovasculares mayores (MACE).

El lector crítico debe preguntarse: ¿Está claramente definida la población (ej. pacientes con y sin enfermedad cardiovascular previa)?, ¿Los desenlaces son clínicamente relevantes o solo incluyen variables intermedias como niveles de glucosa?, ¿La comparación es pertinente para la práctica clínica local? De esta manera, la lectura crítica de la pregunta de investigación permite valorar desde el inicio si la revisión sistemática está bien enfocada y si sus resultados serán aplicables a la toma de decisiones clínicas.

4.2 ELABORACIÓN Y REGISTRO DEL PROTOCOLO

Un criterio fundamental al leer críticamente una revisión sistemática es comprobar si los autores elaboraron y registraron un protocolo previo. Este documento es la garantía de que la revisión se desarrolló con transparencia y rigor metodológico, ya que evita cambios arbitrarios en la pregunta, en los desenlaces o en los métodos de análisis una vez conocidos los resultados. La ausencia de un protocolo suele asociarse con mayor riesgo de sesgo de reporte.

Las guías PRISMA-P y el Cochrane Handbook recomiendan que un protocolo incluya, de forma detallada, la pregunta de investigación, los criterios de inclusión y exclusión de estudios, la estrategia de búsqueda, los métodos para la extracción de datos, la forma de evaluar el riesgo de sesgo y el plan de análisis estadístico. Asimismo, debe declarar potenciales conflictos de interés y describir cómo se resolverán las discrepancias entre revisores.

Existen plataformas reconocidas para su registro, entre ellas PROSPERO (la más utilizada en salud), Open Science Framework (OSF) y los protocolos de revisiones Cochrane. Al leer un artículo, el crítico debe verificar si aparece un número de registro y si lo publicado coincide con lo planificado.

Por ejemplo, en el caso del paciente con diabetes tipo 2 que consulta sobre los inhibidores de SGLT2, supongamos que el médico revisa un metaanálisis registrado en PROSPERO (Nº CRD42021234567). Al compararlo con el artículo final, observa que los desenlaces principales fueron respetados (mortalidad cardiovascular y hospitalización por insuficiencia cardíaca), pero también identifica que se añadió como desenlace primario la reducción de hemoglobina glicosilada, originalmente listado como secundario. Esta modificación puede interpretarse como un cambio post hoc, lo que reduce la confianza en la objetividad de los resultados.

4.3 ESTRATEGIA DE BÚSQUEDA

Otro aspecto clave en la lectura crítica de una revisión sistemática es la estrategia de búsqueda utilizada para identificar los estudios. Una revisión es tan sólida como lo sea su búsqueda: si esta es incompleta o sesgada, los resultados pueden ser engañosos.

Una estrategia rigurosa debe ser amplia, transparente y reproducible. Esto implica detallar las bases de datos consultadas, los términos de búsqueda empleados y los filtros aplicados. Las principales bases utilizadas en salud son PubMed/MEDLINE, Cochrane Library, Embase, Scopus y Web of Science, aunque en Latinoamérica adquieren gran relevancia SciELO y la Biblioteca Virtual en Salud (BVS). Además, se recomienda explorar la literatura gris (tesis, actas de congresos, preprints, registros de ensayos como ClinicalTrials.gov) para reducir el sesgo de publicación.

El uso de términos MeSH (Medical Subject Headings) y DeCS (Descriptores en Ciencias de la Salud), combinados con operadores booleanos (AND, OR, NOT), asegura que la búsqueda capture con precisión la mayor parte de los estudios relevantes. Por ejemplo, un metaanálisis sobre inhibidores de SGLT2 debería mostrar en su sección de métodos las palabras clave utilizadas ("Sodium-GlucoseTransporter 2 Inhibitors", "Diabetes Mellitus, Type 2", "Cardiovascular Outcomes"), así como los límites temporales o de idioma aplicados.

Desde la perspectiva del lector crítico, hay preguntas indispensables: ¿se explicaron claramente las bases de datos consultadas?, ¿se incluyó literatura en distintos idiomas o solo en inglés?, ¿se aportó la estrategia completa de búsqueda como anexo para permitir su reproducción?, ¿se consideraron estudios no publicados o literatura gris? Una revisión que no aclare estos aspectos puede estar sesgada hacia estudios positivos y sobrestimar los beneficios de una intervención.

Volviendo al caso del paciente con diabetes tipo 2, si el médico encuentra una revisión sobre inhibidores de SGLT2 que solo buscó en PubMed y en inglés, debería desconfiar de su exhaustividad. Es posible que estudios relevantes publicados en otras bases o en otros idiomas por ejemplo, en población latinoamericana— hayan quedado fuera, limitando la aplicabilidad de los resultados a su contexto clínico, la forma correcta debería ser descrita de esta forma:

- "Se realizó una búsqueda sistemática en PubMed/MEDLINE, Embase, Cochrane Central Register of Controlled Trials (CENTRAL), Scopus y Web of Science hasta marzo de 2023. Se utilizaron términos MeSH y DeCS relacionados con 'Diabetes Mellitus, Type 2', 'Sodium-GlucoseTransporter 2 Inhibitors', 'Cardiovascular Outcomes', combinados con operadores booleanos. No se aplicaron restricciones de idioma. Además, se consultaron registros de ensayos clínicos (ClinicalTrials.gov) y literatura gris.

4.4 SELECCIÓN DE ESTUDIOS

Una vez realizada la búsqueda bibliográfica, el siguiente paso en una revisión sistemática es la selección de los estudios que serán incluidos en el análisis. Este proceso es crucial porque determina la calidad y la relevancia de la evidencia que finalmente se sintetizará.

Desde la perspectiva de la lectura crítica, lo importante es verificar si los autores describen

con claridad:

Criterios de inclusión y exclusión: deben estar definidos antes de comenzar la selección. Estos criterios suelen basarse en la población, la intervención, el comparador, los desenlaces y el diseño de estudio (por ejemplo, ensayos clínicos aleatorizados para preguntas de eficacia).

Proceso de selección: idealmente, dos revisores deberían evaluar de forma independiente los títulos y resúmenes, para luego revisar el texto completo. Las discrepancias deben resolverse mediante consenso o con un tercer revisor.

Transparencia: la revisión debe mostrar cuántos estudios se identificaron inicialmente, cuántos se excluyeron y por qué motivos. Esto se representa en un diagrama de flujo PRISMA, que facilita al lector entender el proceso.

Si una revisión no especifica los criterios de selección o no explica cómo se resolvieron las discrepancias entre revisores, su validez metodológica se ve comprometida. Del mismo modo, si se excluyeron estudios relevantes sin justificación clara, existe un riesgo de sesgo de selección.

En nuestro ejemplo aplicado al caso clínico, se describe lo siguiente en su metodología:

- “Se identificaron 2.345 referencias tras la búsqueda inicial. Después de eliminar duplicados, 1.876 estudios fueron evaluados por título y resumen. De ellos, 1.820 fueron excluidos por no cumplir los criterios de inclusión (estudios en animales, revisiones narrativas, reportes de casos, población sin diabetes tipo 2). Un total de 56 artículos fueron evaluados en texto completo, y 10 ensayos clínicos aleatorizados cumplieron con los criterios para la síntesis cuantitativa. La selección fue realizada por dos revisores de manera independiente; las discrepancias se resolvieron mediante consenso. El proceso completo se muestra en el diagrama PRISMA del anexo 2.”

Un lector crítico que revisa esta sección puede valorar positivamente la claridad y transparencia de la selección. Sin embargo, si el artículo no incluyera los motivos de exclusión o no presentara un diagrama PRISMA, la confianza en la exhaustividad de la revisión sería menor.

4.5 EXTRACCIÓN DE DATOS

Una vez seleccionados los estudios, el paso siguiente en una revisión sistemática es la extracción de datos. Esta etapa consiste en trasladar de cada artículo los elementos esenciales que permitan compararlos y, posteriormente, sintetizarlos en un análisis narrativo o cuantitativo.

Desde la perspectiva de la lectura crítica, lo importante no es memorizar cómo se llena una tabla de extracción, sino saber identificar si el artículo lo hizo de manera rigurosa y transparente. Una revisión sistemática de calidad siempre presenta una plantilla estandarizada —a veces como tabla en el texto principal, otras en material suplementario— donde resume las características de cada estudio incluido. La información mínima que se espera encontrar comprende:

Características de la población: edad promedio, proporción de hombres y mujeres, comorbilidades, criterios de inclusión y exclusión.

Detalles de la intervención y comparadores: fármaco, dosis, duración del tratamiento o tipo de exposición analizada.

Desenlaces primarios y secundarios: aquellos que guían la pregunta de investigación (mortalidad, eventos cardiovasculares) y los adicionales (biomarcadores, variables intermedias).

Diseño y tamaño muestral: número de participantes, tipo de estudio (ECA, cohorte, etc.) y duración del seguimiento.

Resultados principales: medidas de efecto (RR, OR, HR, medias con desviación estándar) con sus intervalos de confianza.

Riesgo de sesgo y calidad metodológica: evaluación realizada con herramientas reconocidas (RoB 2, ROBINS-I, QUADAS-2, etc.).

Cuando un artículo no detalla esta información, el lector debe ser cauteloso: la falta de transparencia en la extracción de datos puede indicar riesgo de sesgo o incluso errores en la síntesis.

Ejemplo aplicado al caso clínico:

Tabla 2. Características de los estudios incluidos en el metaanálisis sobre inhibidores de SGLT2 en diabetes tipo 2

Estudio (año)	Población	Intervención / Comparador	Desenlaces principales	N (seguimiento)	Resultados	Riesgo de sesgo
Zheng et al., JAMA Cardiol 2020 PMID: 33031522	46.969 pacientes con DM2; edad media 63,7 años; 65,9 % hombres; 66 % con enfermedad CV establecida	iSGLT2 (empagliflozina, dapagliflozina, canagliflozina) vs placebo	Mortalidad CV, MACE, hospitalización por IC, desenlaces renales compuestos	6 ECA; 2-4 años	HR MACE 0,90 (IC95 % 0,85-0,95); HR hospitalización por IC/muerte CV 0,78; HR renal 0,62	Bajo (RoB 2)
Castellana et al., Cardiovasc Diabetol 2020 DOI: 10.1186/s12933-020-01067-8	1.703.519 pacientes de múltiples cohortes; edad media variable; proporción de CV previa dependiente de cohorte	Evaluación de elegibilidad de pacientes para criterios de EMPA-REG, CANVAS, DECLARE-TIMI 58 y VERTIS-CV	Proporción de pacientes que cumplirían criterios de inclusión en CVOTs	5 estudios de cohortes	Elegibilidad: EMPA-REG 17 %, CANVAS 36,4 %, DECLARE 49,5 %, VERTIS-CV 19 %	Moderado (heterogeneidad alta)

Fuente: Elaboración propia.

Cómo interpretarla críticamente en nuestro caso clínico:

- Ver si los artículos incluidos muestran todos los elementos clave (como en la tabla). Si falta alguno (por ejemplo, duración de seguimiento, desenlace renal, subgrupos), eso disminuye la aplicabilidad de los resultados al paciente real.
- Observar si los desenlaces secundarios son relevantes (por ejemplo, reducción de azúcar, peso) pero no sobredimensionarlos frente a los desenlaces clínicos críticos (muerte cardiovascular, hospitalización por ICC).
- Ver el riesgo de sesgo: si los estudios son ensayos aleatorizados, placebo-control, con buen seguimiento, poco desenmascaramiento, y con criterios bien definidos; si no, aplicar reservas al interpretar los resultados.
- El cuadro también muestra que los estudios de mayor tamaño (ej. EMPA-REG, CANVAS) aportan más peso estadístico en el metaanálisis, y que el riesgo de sesgo global fue bajo según la herramienta RoB 2. Para el lector crítico, este nivel de detalle confirma que la extracción de datos fue meticulosa, transparente y útil para interpretar la síntesis de resultados.

4.6 EVALUACIÓN DEL RIESGO DE SESGO Y CALIDAD DE LOS ESTUDIOS

Una revisión sistemática es tan sólida como la calidad de los estudios que la conforman. Por eso, un aspecto esencial en la lectura crítica es comprobar si los autores evaluaron de manera rigurosa el riesgo de sesgo y la calidad metodológica de cada estudio incluido.

El sesgo se entiende como cualquier desviación sistemática que pueda distorsionar la estimación real del efecto de una intervención. La presencia de sesgo en los estudios primarios compromete directamente la validez de los resultados del metaanálisis. Para evitar interpretaciones erróneas, los revisores deben utilizar herramientas estandarizadas, distintas según el tipo de estudio:

- **RoB 2.0 (Risk of Bias 2.0):** para ensayos clínicos aleatorizados. Evalúa aspectos como la generación de la secuencia aleatoria, la ocultación de la asignación, el cegamiento, la integridad de los datos y el reporte selectivo.
- **ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions):** para estudios no aleatorizados, analiza confusión, selección de participantes, clasificación de intervenciones y medición de desenlaces.
- **QUADAS-2:** diseñado para estudios de pruebas diagnósticas, evalúa riesgo de sesgo y aplicabilidad en cuatro dominios (selección de pacientes, prueba índice, estándar de referencia, flujo de pacientes).
- **AMSTAR 2:** orientado a evaluar la calidad de revisiones sistemáticas publicadas, considerando aspectos como la estrategia de búsqueda, la duplicación en selección y extracción, y la evaluación de sesgo en los estudios primarios.

Además de estas herramientas, la certeza global de la evidencia debe valorarse con el sistema GRADE (Grading of Recommendations, Assessment, Development and Evaluation). Este enfoque permite calificar la evidencia por cada desenlace en cuatro niveles (alta, moderada, baja, muy baja), considerando factores como limitaciones metodológicas, inconsistencia,

indirectitud, imprecisión y sesgo de publicación. Los resultados suelen presentarse en una tabla Summary of Findings (SoF), que facilita al lector identificar de manera rápida la calidad de la evidencia y la magnitud del efecto.

Ejemplo aplicado al caso clínico:

Tabla 3. Evaluación del riesgo de sesgo (RoB 2.0) en los principales ensayos clínicos de inhibidores de SGLT2 en diabetes tipo 2

Estudio	Aleatorización	Desviaciones de la intervención	Datos incompletos	Medición de desenlaces	Reporte selectivo	Juicio global
EMPA-REG OUTCOME	Bajo riesgo	Bajo riesgo	Bajo riesgo	Bajo riesgo	Bajo riesgo	Bajo riesgo
CANVAS Program	Bajo riesgo	Algunas preocupaciones	Bajo riesgo	Bajo riesgo	Bajo riesgo	Algunas preocupaciones
DECLARE-TIMI 58	Bajo riesgo	Bajo riesgo	Bajo riesgo	Bajo riesgo	Bajo riesgo	Bajo riesgo

Fuente: Elaboración propia.

Su descripción tendría que ser: “En la revisión sistemática que consulta el médico para responder al paciente con diabetes tipo 2, los autores evaluaron los diez ensayos clínicos incluidos mediante la herramienta RoB 2.0. El resultado fue que la mayoría presentaban bajo riesgo de sesgo, aunque dos estudios mostraron cierta preocupación por la falta de cegamiento de participantes. En el análisis con GRADE, la certeza de la evidencia para el desenlace “mortalidad cardiovascular” fue calificada como alta, mientras que para “hospitalización por insuficiencia cardíaca” se consideró moderada debido a la heterogeneidad entre estudios”.

Para el lector crítico, estos hallazgos permiten interpretar con mayor confianza los resultados del metaanálisis. Una revisión que no aplica herramientas de evaluación de sesgo ni califica la certeza de la evidencia debe considerarse metodológicamente débil, por más sofisticado que sea su análisis estadístico.

4.7 SÍNTESIS DE LA EVIDENCIA

El propósito final de una revisión sistemática es integrar y resumir los hallazgos de los estudios incluidos para responder de manera clara a la pregunta de investigación. Esta síntesis puede adoptar dos formas principales: narrativa o cuantitativa, cada una con sus ventajas y limitaciones.

A) SÍNTESIS NARRATIVA (CUALITATIVA)

Se utiliza cuando los estudios incluidos son demasiado heterogéneos (distintos) entre sí o cuando los desenlaces no son comparables de manera directa. La síntesis narrativa no combina los resultados estadísticamente, sino que describe las tendencias generales, las fortalezas de los estudios, las limitaciones metodológicas y las brechas de conocimiento.

En la lectura crítica, el lector debe valorar si los autores presentan una narrativa equilibrada, sin exagerar resultados positivos ni minimizar los negativos. Además, debe revisar si se discuten de forma explícita las causas de la heterogeneidad (poblaciones distintas, intervenciones con variaciones de dosis, desenlaces definidos de manera desigual).

Este formato es habitual en revisiones narrativas o en scoping reviews, cuyo objetivo es ofrecer un panorama amplio más que una estimación precisa del efecto.

B) METAANÁLISIS (SÍNTESIS CUANTITATIVA)

Cuando los estudios son lo suficientemente homogéneos en diseño, población, intervención y desenlaces, es posible realizar un metaanálisis, que combina estadísticamente los resultados individuales para obtener un estimador global.

Existen dos modelos principales:

- Modelo de efectos fijos: asume que todos los estudios estiman el mismo efecto verdadero, y las diferencias entre ellos se deben únicamente al azar.
- Modelo de efectos aleatorios: considera que los estudios pueden tener efectos verdaderos distintos, por lo que incorpora la variabilidad entre ellos. Este suele ser el modelo más

realista en medicina, donde rara vez los estudios son completamente idénticos.

El metaanálisis permite calcular medidas de efecto como riesgo relativo (RR), odds ratio (OR), hazard ratio (HR), diferencia de medias (MD) o diferencia de medias estandarizada (SMD). Los resultados se presentan con su intervalo de confianza, y la forma gráfica clásica de mostrarlos es el forestplot, que resume visualmente los resultados de cada estudio y la estimación combinada.

Desde la perspectiva crítica, el lector debe preguntarse: ¿Los estudios eran lo suficientemente homogéneos para ser combinados?, ¿El modelo elegido (fijo o aleatorio) fue el más apropiado?, ¿El forestplot muestra resultados consistentes o hay heterogeneidad marcada?, ¿Los autores interpretan los resultados con cautela, o presentan la síntesis como una verdad absoluta?

4.8 ANÁLISIS DE HETEROGENEIDAD Y SESGOS

Al interpretar un metaanálisis, el lector crítico debe recordar que no todos los estudios incluidos son idénticos. Variaciones en la población, la intervención, la forma de medir los desenlaces o la calidad metodológica pueden generar heterogeneidad. Asimismo, la publicación selectiva de resultados positivos puede introducir sesgos de publicación que distorsionen las conclusiones.

Heterogeneidad: La heterogeneidad mide hasta qué punto los resultados de los estudios difieren más de lo esperado por el azar. Una revisión rigurosa debe reportar:

- Prueba de Cochran-Q: evalúa si la variabilidad observada es mayor a la esperada por azar. Su valor p indica significancia, aunque es sensible al número de estudios.
- Estadístico I²: expresa en porcentaje la proporción de variabilidad atribuible a heterogeneidad real
- Estrategias de exploración: análisis de subgrupos y metarregresión, que permiten investigar si la variabilidad se explica por factores como edad, presencia de comorbilidades, duración del seguimiento o dosis de intervención.

Desde la perspectiva crítica, el lector debe revisar si la heterogeneidad fue cuantificada, si se interpretó con cautela y si se justificó adecuadamente el uso de modelos de efectos fijos o aleatorios.

Sesgo de publicación: El sesgo de publicación ocurre cuando los estudios con resultados positivos tienen más probabilidad de ser publicados, mientras que los negativos o neutros permanecen inéditos. Este fenómeno puede llevar a que un metaanálisis sobrestime el efecto de una intervención. Herramientas para detectarlo:

- Gráfico en embudo (funnelplot): representa el tamaño del efecto frente a la precisión del estudio. Un gráfico simétrico sugiere ausencia de sesgo; la asimetría puede indicar estudios faltantes.
- Prueba de Egger: análisis estadístico que cuantifica la asimetría del funnelplot; un $p < 0,05$ sugiere sesgo de publicación.
- Método trim and fill: estima cuántos estudios faltan debido a sesgo de publicación y recalcula la estimación global ajustada.

En lectura crítica, es fundamental revisar si el artículo aplicó estos métodos, cuántos estudios incluyó (el funnelplot y la prueba de Egger no son confiables si hay menos de 10 estudios), y si los autores discutieron las limitaciones relacionadas con el sesgo de publicación. Estas métricas serán analizadas a detalle en el acápite de interpretación estadística.

4.9. OTRAS MODALIDADES DE SÍNTESIS DE EVIDENCIA

Si bien la mayoría de las revisiones sistemáticas y metaanálisis se centran en intervenciones terapéuticas, existen diversas modalidades de síntesis de evidencia que responden a preguntas distintas y que el lector crítico debe aprender a reconocer. Cada una tiene fortalezas y limitaciones que impactan en la confianza de sus conclusiones.

METAANÁLISIS EN RED (NETWORK META-ANALYSIS, NMA)

El NMA permite comparar múltiples tratamientos de manera simultánea, incluso cuando no existen ensayos que enfrenten directamente todas las opciones. Se basa en una red de comparaciones directas e indirectas, bajo el supuesto de que los estudios son suficientemente comparables (transitividad).

Su utilidad radica en que genera rankings de eficacia y seguridad (ej. mediante SUCRA: Surface Underthe Cumulative Ranking).

- Ejemplo: en hipertensión arterial, un NMA puede integrar ensayos que comparan A vs B y B vs C, para inferir la eficacia relativa de A vs C. El lector crítico debe preguntarse si se evaluó la consistencia entre comparaciones directas e indirectas, y si los supuestos de homogeneidad se cumplen.

REVISIONES NARRATIVAS

Son revisiones más tradicionales, que resumen un tema amplio sin seguir una metodología estandarizada. Su fortaleza es ofrecer contexto y actualización general en campos emergentes, pero son muy susceptibles a sesgo de selección, ya que dependen del criterio subjetivo del autor.

- Ejemplo: una revisión narrativa sobre microbiota intestinal y enfermedades autoinmunes puede ser útil para explorar hipótesis, pero no debe utilizarse como base exclusiva para guías de práctica clínica. El lector crítico debe reconocerlas como informativas, pero no concluyentes.

REVISIONES DE ALCANCE (SCOPING REVIEWS)

Exploran un tema amplio sin intentar sintetizar cuantitativamente los resultados. Su objetivo es mapear la literatura existente, identificar vacíos de conocimiento y proponer futuras preguntas de investigación. Se apoyan en guías como PRISMA-ScR.

- Ejemplo: una revisión de alcance podría explorar el uso de inteligencia artificial en diagnóstico radiológico, identificando qué áreas han sido más investigadas y cuáles permanecen inexploradas.

REVISIONES SISTEMÁTICAS CUALITATIVAS

Integran los hallazgos de estudios cualitativos (entrevistas, grupos focales, etnografías). No buscan estimaciones numéricas, sino comprender experiencias, percepciones y contextos de los pacientes.

- Ejemplo: una revisión cualitativa puede analizar las barreras percibidas por los pacientes con diabetes para adherirse a su tratamiento. Para el lector crítico, la clave es valorar si la síntesis cualitativa fue rigurosa, transparente y coherente con la pregunta planteada.

METAANÁLISIS DE PRUEBAS DIAGNÓSTICAS

Evalúan la exactitud diagnóstica de una prueba en comparación con un estándar de referencia. Utilizan modelos bivariados o HSROC (Hierarchical Summary Receiver Operating Characteristic), que permiten estimar sensibilidad, especificidad y curvas resumen ajustadas por heterogeneidad y efecto de umbral.

- Ejemplo: un metaanálisis de troponina ultrasensible para diagnóstico de infarto agudo de miocardio combina estudios con distintos puntos de corte y genera una curva HSROC que permite estimar el rendimiento global de la prueba. En la lectura crítica, se debe verificar si se aplicó QUADAS-2 para evaluar riesgo de sesgo en cada estudio diagnóstico.

4.10 INTERPRETACIÓN Y APLICABILIDAD

El paso final en la lectura crítica de una revisión sistemática o metaanálisis es interpretar los hallazgos y valorar su aplicabilidad en la práctica clínica. No basta con conocer los resultados numéricos; lo esencial es integrarlos al contexto del paciente y del sistema de salud.

En primer lugar, debe analizarse la magnitud y dirección del efecto. Un resultado estadísticamente significativo puede carecer de relevancia clínica si la magnitud del beneficio es pequeña. Aquí cobra importancia el concepto de mínima diferencia clínicamente importante (MCID), que ayuda a determinar si la diferencia observada tiene un impacto real en la vida del paciente.

Otro elemento es la certeza de la evidencia, que debe estar valorada con un sistema estandarizado como GRADE. Una conclusión basada en evidencia de alta certeza ofrece mayor confianza que aquella apoyada en estudios de baja calidad o con alta heterogeneidad.

La relevancia clínica debe evaluarse en función de los desenlaces que importan al paciente: mortalidad, hospitalización, calidad de vida y seguridad del tratamiento. No siempre los resultados en variables intermedias (como biomarcadores) se traducen en beneficios tangibles.

Finalmente, la aplicabilidad depende de factores como el perfil de los pacientes estudiados, el contexto local y el balance entre beneficios y riesgos. Incluso un tratamiento eficaz puede no ser la mejor opción si los efectos adversos, los costos o las barreras de acceso superan a sus ventajas.

En síntesis, la interpretación de una revisión sistemática no se limita a los números: exige un juicio clínico informado, capaz de ponderar la calidad de la evidencia, su relevancia y su aplicabilidad a la realidad del paciente frente al médico.

5. EVALUACIÓN DE LA VALIDEZ Y RELEVANCIA

Al leer una revisión sistemática o un metaanálisis, el primer paso del análisis crítico es preguntarse: ¿es válido lo que estoy leyendo? y, en segundo lugar, ¿es relevante para mis pacientes y mi contexto? Una revisión puede estar impecablemente realizada desde el punto de vista metodológico, pero si no responde a una necesidad clínica concreta o si la población analizada difiere demasiado de la práctica habitual, su utilidad será limitada.

5.1 VALIDEZ INTERNA

La validez interna hace referencia a la credibilidad de los resultados y depende de que la revisión haya seguido una metodología rigurosa y transparente. Para valorarla, el lector debe fijarse en:

- Pregunta claramente definida: si se estructuró en formato PICO o similar.
- Protocolo registrado: idealmente en PROSPERO u OSF, con criterios de inclusión y exclusión preespecificados.
- Búsqueda exhaustiva y reproducible: que abarque varias bases de datos, con uso de términos MeSH/DeCS, sin restricciones arbitrarias de idioma.
- Selección y extracción de datos duplicada: realizada por al menos dos revisores de forma independiente.
- Evaluación sistemática del riesgo de sesgo: con herramientas estandarizadas (RoB 2, ROBINS-I, QUADAS-2, AMSTAR 2).
- Síntesis adecuada: narración clara cuando hay heterogeneidad, metaanálisis cuando los estudios son comparables.
- Análisis de heterogeneidad y sesgos de publicación: reportados con I^2 , Cochran-Q, funnelplot y pruebas estadísticas apropiadas.

Si alguno de estos elementos falta o se reporta de manera superficial, la confianza en los resultados de la revisión disminuye.

5.2 RELEVANCIA CLÍNICA

Una vez valorada la validez interna, el siguiente paso es considerar la relevancia clínica, es decir, si los hallazgos son aplicables al paciente real. Para ello, conviene reflexionar sobre tres aspectos:

- Población: ¿los pacientes incluidos en los estudios se parecen a los que atiendo en mi práctica diaria?
- Desenlaces: ¿los resultados medidos son clínicamente importantes (ej. mortalidad, hospitalización, calidad de vida) o solo variables intermedias?
- Balance riesgo/beneficio y aplicabilidad local: ¿el beneficio supera los riesgos y los costos? ¿el tratamiento está disponible en mi entorno?

Una revisión puede mostrar un beneficio claro, pero si los pacientes estudiados son muy distintos a los de mi contexto, la aplicabilidad será limitada.

6. INTERPRETACIÓN ESTADÍSTICA

La fortaleza de una revisión sistemática o de un metaanálisis no reside únicamente en la calidad de los estudios incluidos ni en la transparencia de la metodología. El verdadero valor surge en el momento de interpretar los resultados estadísticos y comprender lo que significan para la práctica clínica.

Muchos médicos y estudiantes se enfrentan a tablas llenas de siglas (RR, OR, HR, MD, SMD), gráficos con diamantes y cuadrados de distintos tamaños (forestplots), o curvas complejas (HSROC, SUCRA), y sienten que basta con leer la conclusión final del artículo. Sin embargo, la medicina basada en la evidencia exige más que aceptar resúmenes: demanda la capacidad de leer críticamente los números y figuras para responder tres preguntas esenciales:

- ¿Qué mide realmente el resultado estadístico? (ej. un riesgo relativo, una odds ratio o una diferencia de medias).
- ¿Qué tan confiable es esa estimación? (intervalos de confianza, heterogeneidad, sesgo de publicación).
- ¿Qué significa para mi paciente? (certeza de la evidencia, mínima diferencia clínicamente importante, aplicabilidad).

La interpretación estadística, por tanto, es el puente entre los resultados de la investigación y la toma de decisiones clínicas informadas. Sin ella, se corre el riesgo de basarse en conclusiones incompletas o incluso engañosas.

6.1 INTERPRETACIÓN DE LAS REVISIONES NARRATIVAS

Las revisiones narrativas son el formato más común en la literatura médica tradicional. A diferencia de una revisión sistemática, no siempre siguen un protocolo riguroso ni aplican criterios de búsqueda exhaustivos. Suelen tener un carácter más descriptivo y explicativo, reuniendo estudios seleccionados por el autor para ofrecer una visión general sobre un tema.

Desde la perspectiva de la lectura crítica, lo fundamental es entender que la interpretación de una revisión narrativa depende casi por completo de la coherencia argumentativa del autor y de la amplitud de la literatura considerada.

Aspectos clave que debe analizar el lector crítico:

Alcance y objetivos: ¿la revisión está orientada a un área concreta o se limita a un resumen superficial?

Selección de la literatura: ¿el autor explica cómo eligió los artículos o solo cita los que apoyan su postura?

Equilibrio en la presentación: ¿se incluyen tanto estudios positivos como negativos, o se privilegian resultados favorables?

Identificación de limitaciones: ¿el autor señala las brechas de conocimiento y los posibles sesgos, o presenta sus conclusiones como verdades absolutas?

La lectura crítica de una revisión narrativa debe enfocarse en extraer las tendencias generales: qué aspectos parecen más consistentes en la literatura y cuáles son aún inciertos o contradictorios. No se debe asumir que una conclusión es definitiva, sino considerarla como un punto de partida para el razonamiento clínico o la generación de nuevas preguntas de investigación.

6.2 INTERPRETACIÓN DE UN METAANÁLISIS CLÁSICO

Cuando los estudios primarios son comparables en diseño, población y desenlaces, los autores pueden realizar un metaanálisis, es decir, una síntesis estadística que combina cuantitativamente sus resultados. Para el lector crítico, el reto es comprender qué significa el estimador global y cómo interpretarlo clínicamente.

Medidas de efecto: Las más utilizadas en medicina son:

- Riesgo relativo (RR): indica cuántas veces es más probable que ocurra un evento en el grupo intervención respecto al comparador.
- Odds ratio (OR): compara las probabilidades relativas; más común en estudios de casos y controles.
- Hazard ratio (HR): se usa en análisis de supervivencia, donde importa no solo si ocurre el evento, sino cuándo.
- Diferencia de medias (MD): compara directamente medias entre dos grupos (ej. reducción en mmHg de presión arterial).

- Diferencia de medias estandarizada (SMD): se utiliza cuando los estudios miden el mismo desenlace con escalas diferentes (ej. dolor en diferentes escalas).

Tabla 4. Medidas de efecto en metaanálisis clásicos

Medida	Uso principal	Valor de no efecto	Interpretación crítica
Riesgo relativo (RR)	Ensayos clínicos; cohorte	1	RR < 1: reducción de riesgo; RR > 1: aumento de riesgo
Odds ratio (OR)	Casos y controles	1	OR < 1: menor probabilidad; OR > 1: mayor probabilidad
Hazard ratio (HR)	Estudios de supervivencia	1	HR < 1: menor riesgo en el tiempo; HR > 1: mayor riesgo
Diferencia de medias (MD)	Comparaciones directas en misma escala	0	MD ≠ 0: diferencia en unidades reales (ej. mmHg, kg)
Diferencia de medias estandarizada SMD	Comparaciones con distintas escalas	0	Estándar: 0.2 pequeño, 0.5 moderado, 0.8 grande (Cohen)

Fuente: Elaboración propia.

Intervalos de confianza: El IC 95 % no solo indica significación estadística, sino también precisión. Intervalos estrechos refieren estimación confiable. Intervalos amplios refieren gran incertidumbre.

Para comprender su interpretación presentamos la siguiente tabla:

Tabla 5. Efecto de los inhibidores SGLT2 en DM2 (síntesis ficticia para fines didácticos)

Desenlace	Medida	Estimación (IC95%)	Valor de no efecto	Modelo	Estudios (n)	I ² (%)
MACE (ictus/IM/muerte CV)	RR	0.90 (0.85–0.95)	1.00	Aleatorio	10	18
Hospitalización por IC	HR	0.78 (0.72–0.84)	1.00	Aleatorio	10	42
Mortalidad por cualquier causa	OR	0.92 (0.86–0.99)	1.00	Aleatorio	10	20
HbA1c (%)	MD	–0.35 (–0.45 a –0.25)	0.00	Aleatorio	9	30
Peso corporal (kg)	MD	–2.1 (–2.8 a –1.4)	0.00	Aleatorio	8	36
PAS (mmHg)	MD	–3.5 (–5.0 a –2.0)	0.00	Aleatorio	7	28
Calidad de vida (escala variable)	SMD	0.20 (0.10–0.31)	0.00	Aleatorio	5	24

Notas: Datos simulados con fines pedagógicos. RR: riesgo relativo; OR: odds ratio; HR: hazard ratio; MD: diferencia de medias; SMD: diferencia de medias estandarizada; IC95%: intervalo de confianza del 95%; PAS: presión arterial sistólica; IM: infarto de miocardio; CV: cardiovascular; IC: insuficiencia cardíaca.

Para su interpretación: el riesgo relativo (RR) para eventos cardiovasculares mayores (MACE) fue de 0.90 (IC95% 0.85–0.95). Esto significa que los pacientes tratados con estos fármacos tuvieron un 10 % menos de riesgo relativo de presentar un MACE en comparación con el grupo control. Como el intervalo de confianza no cruza la unidad, el resultado es estadísticamente significativo y, además, clínicamente relevante, pues la reducción es consistente y se observa con baja heterogeneidad.

En cuanto a la hospitalización por insuficiencia cardíaca, el análisis mostró un hazard ratio (HR) de 0.78 (IC95% 0.72–0.84). A diferencia del RR, el HR incorpora la dimensión temporal, de modo que este resultado indica que los iSGLT2 no solo reducen la incidencia de hospitalizaciones, sino que también retrasan su aparición en el tiempo. El beneficio es claro y robusto, con una reducción relativa del 22 %.

Respecto a la mortalidad por cualquier causa, la medida utilizada fue el odds ratio (OR), que resultó en 0.92 (IC95% 0.86–0.99). Este hallazgo refleja una reducción modesta en la probabilidad de muerte. Aunque estadísticamente significativo, el límite inferior del intervalo de confianza se acerca mucho a la unidad, lo que invita a interpretar el efecto con cautela, pues la magnitud del beneficio es pequeña y su relevancia clínica discutible.

Para los desenlaces intermedios se emplearon medidas de diferencia de medias. En el caso de la hemoglobina glicosilada, la diferencia de medias (MD) fue de –0.35 % (IC95% –0.45 a –0.25), lo que implica una reducción modesta pero consistente en el control glucémico. En el peso corporal, la MD fue de –2.1 kg (IC95% –2.8 a –1.4), un efecto clínicamente más significativo, pues la pérdida de alrededor de dos kilogramos puede impactar en el control metabólico y cardiovascular a largo plazo.

Finalmente, en desenlaces más subjetivos como la calidad de vida se utilizó la diferencia de medias estandarizada (SMD), que fue de 0.20 (IC95% 0.10–0.31). Este valor corresponde a un efecto pequeño según los criterios convencionales, pero relevante al indicar que los pacientes tratados con iSGLT2 reportaron una mejoría consistente en su percepción de salud.

En resumen, la interpretación de estas medidas depende siempre de dos elementos: el valor puntual de la estimación y su intervalo de confianza. En RR, OR y HR, valores por debajo de 1 favorecen a la intervención; en MD y SMD, valores por debajo de 0 indican reducción y por encima de 0, incremento. Sin embargo, la verdadera lectura crítica consiste en valorar no solo la significación estadística, sino también la magnitud del efecto y su pertinencia clínica en relación con el paciente real.

HETEROGENEIDAD

Uno de los errores más frecuentes al leer un metaanálisis es asumir que la estimación global representa un efecto único y aplicable a todos los contextos. En realidad, los estudios incluidos casi nunca son idénticos: difieren en características de los pacientes, intervenciones, dosis, desenlaces medidos o duración del seguimiento. Esta variabilidad se conoce como heterogeneidad y puede modificar de forma importante la interpretación de los resultados.

Existen diferentes tipos de heterogeneidad:

- Clínica: se refiere a diferencias en la población (edad, comorbilidades), en la intervención (fármacos, dosis, duración), en el comparador o en los desenlaces evaluados.
- Metodológica: proviene de diferencias en el diseño de los estudios (ensayos ciegos vs. abierto, tamaños muestrales, criterios de inclusión y exclusión).
- Estadística: es la variabilidad numérica observada en los resultados de los estudios, que se mide con pruebas específicas.

El lector crítico debe tener presente que la heterogeneidad clínica y metodológica son inevitables en investigación, pero lo importante es valorar si estas diferencias son lo suficientemente grandes como para invalidar la síntesis cuantitativa.

Para la medición de la heterogeneidad se utilizan las siguientes medidas estadísticas:

Prueba de Cochran-Q: Es una prueba de hipótesis que evalúa si la variabilidad entre los estudios es mayor de lo que cabría esperar por azar. Un valor de $p < 0,05$ indica heterogeneidad significativa. Sin embargo, tiene limitaciones: cuando hay pocos estudios puede no detectar diferencias reales, y cuando hay muchos estudios tiende a detectar heterogeneidad incluso si esta es clínicamente irrelevante.

Estadístico I^2 : Es el indicador más utilizado en la práctica. Expresa en porcentaje la proporción de la variabilidad total que se debe a heterogeneidad real y no al azar, para su interpretación:

- 0–25 %: heterogeneidad baja.
- 25–50 %: heterogeneidad moderada.
- 50–75 %: heterogeneidad alta.
- 75 %: heterogeneidad muy alta.

Aunque estas categorías son orientativas, el juicio clínico es clave: un I^2 del 40 % puede ser aceptable si los estudios son consistentes en la dirección del efecto, mientras que un I^2 bajo no garantiza ausencia de problemas si los estudios difieren en calidad o población.

ESTRATEGIAS PARA EXPLORAR LA HETEROGENEIDAD

Cuando se detecta heterogeneidad, el metaanálisis no debe detenerse en el reporte del I^2 , sino explorar sus causas. Las principales estrategias son:

- Análisis de subgrupos: divide los estudios según una característica específica (ej. pacientes con vs. sin enfermedad cardiovascular previa, dosis alta vs. baja del fármaco) y analiza si el efecto cambia entre subgrupos.
- Metarregresión: técnica estadística más avanzada que permite evaluar la relación entre características de los estudios (ej. edad media de los pacientes, duración del seguimiento) y el tamaño del efecto.
- Análisis de sensibilidad: repite el metaanálisis excluyendo ciertos estudios (ej. aquellos con alto riesgo de sesgo o tamaños extremos) para comprobar si los resultados globales cambian.

La heterogeneidad no invalida automáticamente un metaanálisis, pero obliga a interpretarlo con cautela. Lo importante no es solo el número (I^2 o p del Q), sino si existe una explicación clínica

plausible para esa variabilidad y si los autores exploraron adecuadamente sus causas.

FOREST PLOT

El forestplot (o gráfico de bosque) es la representación visual más utilizada para resumir los resultados de un metaanálisis. Su valor radica en que, de un solo vistazo, permite observar los efectos de los estudios individuales y la estimación global combinada.

Los elementos principales de un forestplot son:

- Estudios individuales: cada ensayo aparece representado por un cuadrado en el eje horizontal. La posición del cuadrado indica el valor del estimador puntual (RR, OR, HR, MD o SMD), mientras que la línea horizontal que lo acompaña corresponde a su intervalo de confianza (IC 95 %).
 - Tamaño del cuadrado: refleja el peso estadístico del estudio dentro del metaanálisis, que depende sobre todo del tamaño de la muestra y de la precisión de sus resultados. Un ensayo grande como EMPA-REG OUTCOME suele tener un cuadrado más grande que un estudio pequeño.
 - Línea vertical de no efecto: situada en el valor 1 cuando se representan RR, OR o HR, o en 0 cuando se representan diferencias de medias. Si la línea de IC de un estudio cruza esta línea de no efecto, se interpreta que ese estudio por sí solo no tiene un resultado estadísticamente significativo.
 - Diamante final: representa la estimación global del metaanálisis. Su centro indica el valor combinado del efecto y sus extremos el intervalo de confianza. Si el diamante no cruza la línea de no efecto, el resultado global es estadísticamente significativo.
 - Medidas de heterogeneidad: habitualmente se muestran en una esquina del gráfico (I^2 , Cochran-Q), para que el lector pueda valorar la consistencia entre estudios.
- Así mismo para interpretar un Forest plot se presentan los siguientes tips:
- Observar la dirección de los efectos individuales: si la mayoría de los estudios apuntan hacia el mismo lado del gráfico, la conclusión será más robusta.
 - Revisar la magnitud de los efectos: aunque todos vayan en la misma dirección, algunos pueden mostrar beneficios grandes y otros más modestos.
 - Valorar los intervalos de confianza: si son amplios, los resultados son menos precisos; si son estrechos, más confiables.
 - Identificar el peso relativo de cada estudio: un ensayo de gran tamaño puede dominar la estimación global; el lector debe ser consciente de ello.
 - Analizar el diamante global: ¿cruza o no la línea de no efecto?, ¿su IC es estrecho o amplio?
 - Relacionar con la heterogeneidad: incluso si el diamante muestra un beneficio, una alta heterogeneidad reduce la certeza del resultado.

Para su correcta interpretación se presentan las siguiente figura:

Figura 1: Forest plot de eventos cardiovasculares mayores (MACE) con inhibidores de SGLT2 (desenlace dicotómico, medida = RR)

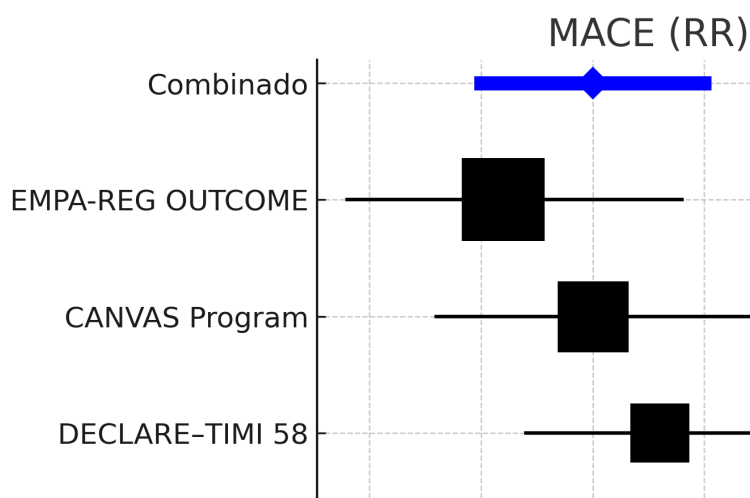


Figura 1. Forest plot de eventos cardiovasculares mayores (MACE) con inhibidores de SGLT2.

El resultado global mostró un RR = 0.90 (IC95% 0.85–0.95). (5) Heterogeneidad: $I^2 = 18\%$, $p = 0.21$.

Para su interpretación de estudios individuales:

- EMPA-REG OUTCOME: RR ≈ 0.86 con IC95% 0.79–0.94. No cruza la línea de no efecto $\rightarrow$ resultado significativo, con un beneficio claro.
- CANVAS Program: RR ≈ 0.90 (IC95% 0.83–0.98). También significativo, aunque con un efecto más cercano a 1.
- DECLARE-TIMI 58: RR ≈ 0.93 (IC95% 0.87–1.00). Resultado en el límite; el IC toca la línea de no efecto, lo que indica un beneficio modesto y con menor certeza.
- VERTIS-CV: RR ≈ 0.97 (IC95% 0.89–1.06). El intervalo cruza la línea de no efecto $\rightarrow$ no significativo.

Para la estimación combinada:

El diamante resume el efecto global: RR = 0.90 (IC95% 0.85–0.95), el centro del diamante indica la estimación puntual (10 % de reducción en riesgo relativo), los extremos corresponden al IC95% que no cruza la línea de no efecto, el beneficio global es estadísticamente significativo.

Para la heterogeneidad:

El análisis reporta $I^2 = 18\%$, $p = 0.21$, lo que indica baja heterogeneidad. Esto significa que los estudios son consistentes entre sí y que las diferencias observadas probablemente se deban al azar más que a variaciones clínicas importantes.

Como interpretación final: Este forestplot muestra que los inhibidores de SGLT2 reducen el riesgo de eventos cardiovasculares mayores (MACE) en un 10 % relativo, con un intervalo de confianza estrecho y baja heterogeneidad. Aunque algunos estudios aislados no alcanzan significación (DECLARE, VERTIS), la dirección del efecto es consistente y la síntesis global es robusta.

Figura 2. Forest plot de cambio en HbA1c (%) con inhibidores de SGLT2 (desenlace continuo, medida = diferencia de medias)

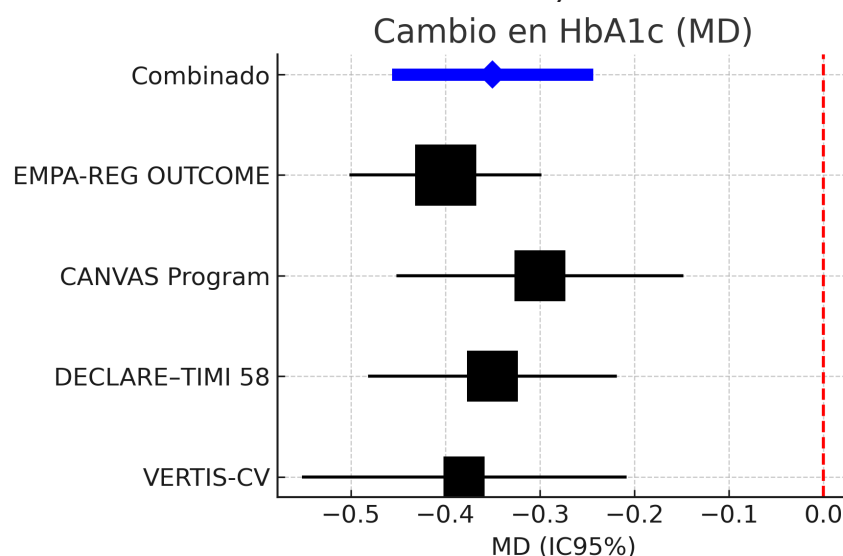


Figura 2. Forest plot de cambio en HbA1c (%) con inhibidores de SGLT2. El resultado global mostró una diferencia de medias (MD) = -0.35 % (IC95% -0.45 a -0.25). Heterogeneidad: $I^2 = 30\%$, $p = 0.15$.

Para la interpretación de los estudios individuales:

- EMPA-REG OUTCOME: MD $\approx -0.40\%$ (IC95% -0.50 a -0.30). Reducción clara y significativa de HbA1c.
- CANVAS Program: MD $\approx -0.30\%$ (IC95% -0.45 a -0.15). También reducción significativa.
- DECLARE-TIMI 58: MD $\approx -0.35\%$ (IC95% -0.48 a -0.22). Efecto consistente.
- VERTIS-CV: MD $\approx -0.38\%$ (IC95% -0.55 a -0.21). Beneficio significativo.

Para la interpretación de estudios combinados:

- El diamante que representa la diferencia de medias combinada, cuenta con MD = -0.35 % (IC95% -0.45 a -0.25), como no cruza la línea de no efecto, el resultado es estadísticamente significativo, la magnitud del efecto es modesta (una reducción de HbA1c de 0.35 %), aunque clínicamente puede ser relevante si se suma a otros beneficios cardiovasculares y renales de estos fármacos.

Para el análisis de heterogeneidad:

- El análisis reporta $I^2 = 30 \%$, $p = 0.15$, lo que corresponde a heterogeneidad moderada. Esto indica que hay cierta variabilidad entre estudios, pero no lo suficiente para invalidar la síntesis. La consistencia general es buena.

El forestplot muestra que los inhibidores de SGLT2 reducen la HbA1c en un promedio de 0.35 puntos porcentuales respecto al comparador, con un intervalo de confianza estrecho y resultados consistentes entre estudios. Aunque la magnitud del beneficio es estadísticamente significativa y clínicamente modesta, cobra importancia cuando se interpreta junto con otros efectos metabólicos (pérdida de peso, reducción de presión arterial) y cardiovasculares.

FUNNELPLOT

El Funnelplot (o gráfico en embudo) es una herramienta visual diseñada para detectar posibles sesgos de publicación en un metaanálisis. Su lógica es sencilla: si todos los estudios realizados sobre un tema se publicaran, sus resultados deberían distribuirse de manera equilibrada alrededor del efecto global. Sin embargo, cuando los estudios pequeños con resultados negativos o neutros no llegan a publicarse, la distribución se vuelve asimétrica.

Para que se entienda mejor, el sesgo de publicación ocurre cuando los resultados de los estudios influyen en su probabilidad de ser publicados. En la práctica, los estudios con resultados positivos (es decir, aquellos que muestran un beneficio claro de la intervención) tienen más posibilidades de ser aceptados en revistas científicas, mientras que los estudios con resultados negativos o neutros suelen publicarse con menor frecuencia o incluso no llegar a difundirse.

Esto genera un problema importante: si los lectores solo tienen acceso a los estudios publicados, la impresión general puede ser que una intervención es más eficaz o más segura de lo que realmente es. En otras palabras, la evidencia disponible está sesgada hacia los resultados favorables.

Los estudios negativos no deben interpretarse como fracasos, sino como piezas fundamentales para conocer el verdadero efecto de una intervención. Un metaanálisis sin ellos puede sobreestimar los beneficios y pasar por alto riesgos relevantes.

El funnelplot busca precisamente detectar este sesgo. Si los estudios negativos pequeños no se publican, el gráfico pierde simetría y aparece un “hueco” en uno de los lados del embudo. Esa asimetría alerta al lector crítico de que la evidencia presentada podría no reflejar toda la realidad, sino solo la parte más favorable.

Las características principales de este gráfico son:

Ejes:

- El eje horizontal representa el estimador de efecto (RR, OR, HR o MD).
- El eje vertical muestra la precisión del estudio, generalmente expresada como la inversa del error estándar ($1/SE$). Así, los estudios grandes y precisos aparecen en la parte superior, mientras que los estudios pequeños y menos precisos quedan en la parte inferior.
- Línea central: corresponde al efecto global estimado por el metaanálisis.
- Líneas punteadas en forma de embudo: delimitan el intervalo de confianza esperado (95 %) alrededor del efecto global. Si no hay sesgo, los estudios deberían agruparse dentro de este embudo de manera simétrica.

Simetría vs. asimetría:

- Simetría: sugiere que no hay evidencia de sesgo de publicación; tanto estudios positivos como neutros aparecen distribuidos a ambos lados del efecto global.
- Asimetría: cuando se observa un “hueco” en uno de los lados, suele indicar que faltan estudios (generalmente los pequeños con resultados negativos), lo cual puede sobreestimar el efecto global.

El lector crítico debe recordar que igual el Funnelplot tiene limitaciones:

- Solo es confiable cuando hay al menos 10 estudios; con menos, la asimetría puede ser producto del azar.
- La asimetría no siempre se debe a sesgo de publicación: también puede reflejar heterogeneidad clínica, diferencias metodológicas o variaciones en la calidad de los estudios.
- Para cuantificar formalmente la asimetría se emplean pruebas estadísticas como la prueba de Egger.

Prueba de Egger: Además de la inspección visual del funnelplot, existen métodos estadísticos para evaluar formalmente la presencia de sesgo de publicación. El más utilizado es la prueba de Egger, que analiza la relación entre el tamaño del efecto y su precisión.

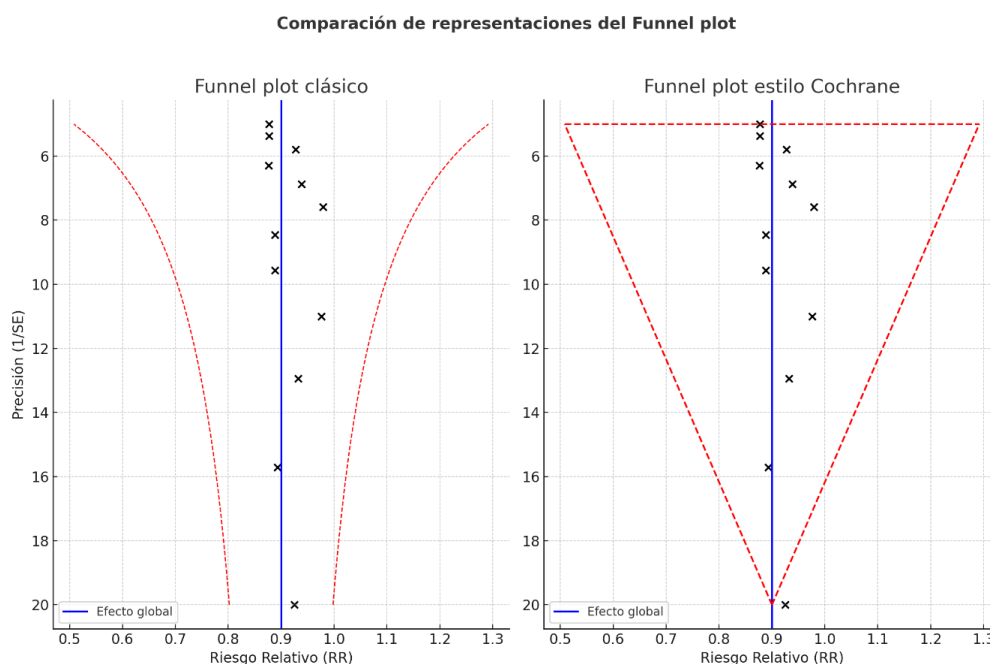
En términos simples, la prueba evalúa si los estudios pequeños tienden a mostrar resultados

diferentes de los estudios grandes. Cuando esto ocurre, puede ser señal de que los estudios pequeños con resultados negativos no fueron publicados.

Para su interpretación: un valor de $p < 0.05$ en la prueba de Egger sugiere asimetría significativa en el funnelplot y, por tanto, posible sesgo de publicación. Sus principales limitaciones radican que, al igual que el funnelplot, la prueba de Egger no es confiable si hay menos de 10 estudios en el metaanálisis, además, la asimetría detectada no siempre se debe a sesgo de publicación, pues también puede reflejar heterogeneidad u otros factores metodológicos.

Con todo esto podemos interpretar el siguiente Funnelplot:

Figura 3. Representaciones del funnelplot para evaluar sesgo de publicación en metaanálisis



Para su interpretación:

En este gráfico cada punto representa un estudio sobre inhibidores de SGLT2 y eventos cardiovasculares mayores (datos ficticios para fines didácticos). Los estudios grandes y más precisos aparecen en la parte superior, muy cercanos al efecto global (línea azul), mientras que los estudios pequeños, menos precisos, se ubican en la parte inferior y presentan mayor dispersión.

La distribución observada es simétrica a ambos lados de la línea de efecto global (RR = 0.90). Esta simetría indica que los estudios publicados incluyen tanto resultados más favorables como menos favorables, pero en conjunto forman la figura de un embudo regular.

Las líneas rojas punteadas delimitan el intervalo esperado del 95 %. La mayoría de los estudios se concentran dentro de esta área, lo que refuerza la impresión de que no existe un sesgo de publicación evidente.

En términos prácticos, no parece que falten “estudios negativos” en un lado del embudo. De existir sesgo de publicación, el gráfico se vería asimétrico, con una acumulación de estudios pequeños solo en el lado de los resultados positivos.

Finalmente, la prueba de Egger aplicada a estos datos simulados fue no significativa ($p = 0.28$), lo que coincide con la inspección visual: no se detecta asimetría relevante y, por tanto, no hay evidencia de sesgo de publicación.

Conclusión: en este ejemplo, el funnelplot da confianza al lector crítico de que el efecto global del metaanálisis (RR = 0.90 para MACE) no está distorsionado por ausencia sistemática de estudios neutros o negativos.

ANÁLISIS DE SENSIBILIDAD

El análisis de sensibilidad es un paso esencial en todo metaanálisis porque permite responder a una pregunta clave: ¿son las conclusiones estables o dependen de ciertas decisiones metodológicas? En otras palabras, nos ayuda a saber si el resultado es robusto. Si los resultados

se mantienen estables ante cambios metodológicos plausibles, la confianza del lector aumenta; si cambian de forma relevante, las conclusiones deben matizarse.

DIFERENCIAS ENTRE SUBGRUPOS, META REGRESIÓN Y ANÁLISIS DE SENSIBILIDAD

En la lectura crítica de un metaanálisis es común encontrar tres tipos de análisis complementarios: subgrupos, metarregresión y sensibilidad. Aunque a veces se mencionan juntos, cumplen propósitos distintos y conviene diferenciarlos claramente.

Los análisis de subgrupos dividen los estudios según una característica clínica o metodológica y recalculan el efecto dentro de cada grupo. Por ejemplo, en nuestro estudio puede presentar un análisis separado para pacientes con enfermedad cardiovascular establecida frente a aquellos sin antecedentes. El objetivo aquí es explorar si la magnitud del beneficio varía en función de una característica relevante de la población, la intervención o el diseño.

La **metarregresión** va un paso más allá: en lugar de dividir en categorías, utiliza un modelo estadístico para evaluar si una variable continua (edad media, duración del seguimiento, proporción de mujeres, dosis del fármaco, etc.) se asocia con la magnitud del efecto. Es una técnica poderosa, pero requiere un número elevado de estudios y puede generar conclusiones espurias si se aplica de manera inadecuada.

El **análisis de sensibilidad**, en cambio, tiene un propósito diferente. No busca explicar por qué los estudios son distintos, sino comprobar si las conclusiones del metaanálisis son estables cuando se cambian ciertas decisiones analíticas. Aquí no se trata de explorar heterogeneidad clínica, sino de evaluar la robustez metodológica. Por ejemplo, ¿qué pasa si excluimos los estudios de menor calidad? ¿O si usamos un modelo de efectos fijos en lugar de aleatorios? ¿O si eliminamos un ensayo con resultados extremos? Si las conclusiones se mantienen en todos estos escenarios, el lector puede confiar más en el resultado.

En resumen, los subgrupos y la metarregresión son herramientas para comprender la variabilidad entre estudios, mientras que el análisis de sensibilidad es una prueba de resistencia: pone a los resultados bajo diferentes condiciones para ver si siguen en pie.

Estrategias habituales de análisis de sensibilidad

El lector crítico debe reconocer que no existe una única forma de hacer sensibilidad. Lo importante es comprobar si el resultado se mantiene estable cuando se cambian ciertos supuestos. A continuación, se describen las estrategias más frecuentes, con ejemplos aplicados al caso clínico de inhibidores de SGLT2.

Exclusión por calidad o riesgo de sesgo: Una de las formas más comunes consiste en repetir el metaanálisis excluyendo los estudios con alto riesgo de sesgo. Por ejemplo, si se elimina un ensayo abierto o con pérdidas elevadas en el seguimiento y el resultado se mantiene, la conclusión es más confiable. Si, en cambio, el efecto desaparece al excluir estudios de baja calidad, la evidencia es frágil.

Análisis “leave-one-out” (influencia de un estudio): Esta técnica elimina un estudio por vez y recalcula el estimador global. Sirve para saber si un ensayo específico está “dominando” el resultado. En el caso de los iSGLT2, si al retirar EMPA-REG OUTCOME (uno de los estudios más grandes) el beneficio desapareciera, la robustez sería cuestionable. En nuestro ejemplo ficticio, quitar cualquier estudio apenas modifica la estimación, lo cual demuestra estabilidad.

Modelos alternativos: efectos fijos vs aleatorios: El modelo de efectos fijos supone que todos los estudios estiman el mismo efecto verdadero; el de efectos aleatorios asume que puede haber variabilidad real entre ellos. Comparar ambos es una prueba de consistencia: si el beneficio solo se observa con un modelo y desaparece con el otro, la certeza es baja. En general, el lector debería desconfiar de resultados que dependen exclusivamente de un modelo.

Medidas de efecto alternativas: En ocasiones, el desenlace puede expresarse con diferentes métricas. Por ejemplo, en eventos poco frecuentes es más apropiado el odds ratio (OR) que el riesgo relativo (RR). Un análisis de sensibilidad puede recalcularse usando medidas alternativas, o aplicar métodos de varianza distintos (como Hartung-Knapp). Si las conclusiones se mantienen, la evidencia gana solidez.

Decisiones sobre manejo de datos: No todos los artículos reportan sus datos de la misma forma, y los autores deben tomar decisiones sobre cómo analizarlos. Ejemplos incluyen el uso de población intention-to-treat vs. per-protocol, imputaciones para valores perdidos o la corrección de continuidad en celdas con cero eventos. El lector crítico debe fijarse si esas decisiones alteran o no los resultados.

Definiciones de desenlace: Los desenlaces pueden variar según la definición empleada.

En estudios cardiovasculares, por ejemplo, “MACE” puede definirse como la combinación de muerte cardiovascular, infarto y accidente cerebrovascular, o incluir también hospitalización por angina. Un análisis de sensibilidad compara estas definiciones más estrictas o más amplias. Si los resultados cambian de manera importante, las conclusiones deben interpretarse con cautela.

Horizonte temporal: Algunos metaanálisis presentan desenlaces a corto y a largo plazo. Comparar seguimientos ≥ 24 meses con seguimientos más breves puede mostrar si el beneficio es sostenido o solo transitorio. Por ejemplo, un fármaco podría reducir hospitalizaciones en los primeros meses pero no mantener el efecto a largo plazo.

Tamaño de estudio: Los estudios pequeños suelen mostrar efectos más extremos (positivos o negativos) y son más susceptibles de sesgo. Un análisis de sensibilidad puede excluirlos para verificar si los estudios grandes sostienen la conclusión. Si el beneficio desaparece al considerar solo los ensayos más grandes, la confianza debe disminuir.

Ajuste por sesgo de publicación: Métodos como el trim-and-fill permiten estimar cuántos estudios “faltantes” serían necesarios para equilibrar el funnelplot y cómo se modificaría el efecto global si se incorporaran. Aunque es un método imperfecto, ofrece una idea de la magnitud del sesgo de publicación potencial.

Metaanálisis acumulativo (cumulative o chronological MA): Esta técnica consiste en añadir los estudios en orden temporal y recalcular el efecto cada vez. Permite visualizar cómo ha evolucionado la evidencia a lo largo del tiempo. Si el efecto se mantiene estable a medida que se agregan más ensayos, la conclusión es más sólida; si varía drásticamente, sugiere fragilidad de la evidencia.

En síntesis: todas estas estrategias son como “golpes de estrés” aplicados al metaanálisis. Si los resultados permanecen estables, podemos confiar en que la conclusión es robusta. Si cambian demasiado, es señal de que la evidencia es más frágil de lo que aparenta.

Tabla 6. Estrategias de análisis de sensibilidad en metaanálisis

Estrategia	Objetivo principal	Interpretación si cambia el resultado
Exclusión por calidad/riesgo de sesgo	Evaluar si los resultados dependen de estudios con alto riesgo de sesgo.	Si el efecto desaparece → la conclusión depende de estudios poco confiables.
Leave-one-out (influencia)	Comprobar si un estudio individual “domina” la síntesis.	Si al quitar un ensayo el resultado cambia drásticamente → el efecto no es robusto.
Modelos alternativos (fijo vs aleatorio)	Verificar consistencia entre supuestos estadísticos.	Si solo un modelo muestra beneficio → certeza baja.
Medida de efecto alternativa	Confirmar que el efecto se mantiene con distintas métricas (RR, OR, HR, MD).	Si la dirección cambia según la medida → interpretar con cautela.
Manejo de datos (ITT vs per-protocol, imputaciones, ceros)	Evaluar impacto de decisiones analíticas.	Si el efecto depende de una sola forma de análisis → baja robustez.
Definiciones de desenlace (estricto vs amplio)	Verificar consistencia según cómo se defina el desenlace.	Si las definiciones cambian el resultado → las conclusiones son sensibles a la definición usada.
Horizonte temporal (corto vs largo plazo)	Evaluar estabilidad del efecto en el tiempo.	Si el beneficio desaparece en seguimientos prolongados → evidencia limitada.
Tamaño de estudio (grandes vs pequeños)	Comprobar si los estudios grandes sostienen la conclusión.	Si el efecto depende de estudios pequeños → riesgo de sobreestimación.
Ajuste por sesgo de publicación (trim-and-fill)	Estimar impacto de estudios faltantes.	Si el efecto ajustado pierde significación → posible sesgo de publicación.
Metaanálisis acumulativo (chronological)	Ver la evolución del efecto a medida que se acumula evidencia.	Si el efecto es inestable con nuevos estudios → conclusión frágil.

Fuente: Elaboración propia.

6.3 INTERPRETACIÓN DE METAANÁLISIS DE RED

El metaanálisis en red, conocido por sus siglas en inglés como NMA (Network Meta-Analysis), representa una evolución del metaanálisis tradicional. Mientras que en un metaanálisis clásico solo se pueden comparar dos intervenciones a la vez —por ejemplo, un fármaco frente a placebo—, el NMA permite integrar de manera simultánea múltiples tratamientos dentro de un mismo análisis. Esto resulta especialmente valioso en áreas de la medicina donde existen

numerosas opciones terapéuticas y no todos los comparadores han sido enfrentados entre sí en ensayos clínicos directos.

La lógica que sostiene este enfoque es sencilla: si un conjunto de ensayos compara el tratamiento A contra el tratamiento B, y otro conjunto compara B contra C, es posible inferir indirectamente la comparación entre A y C. De esta manera, el NMA combina tanto la evidencia directa (cuando los ensayos comparan intervenciones cara a cara) como la evidencia indirecta (cuando la comparación se establece a través de un tercer tratamiento común).

GRÁFICOS DE RED O NETWORK PLOT

El elemento visual más característico de un metaanálisis en red es el gráfico de red o networkplot. En este diagrama, cada intervención se representa como un nodo, y las líneas que conectan a los nodos reflejan la existencia de comparaciones directas entre esos tratamientos. Cuanto más gruesa es la línea, mayor es el número de estudios que sustentan esa comparación.

La interpretación de este gráfico es fundamental porque ofrece una visión panorámica de la evidencia disponible. Un nodo conectado únicamente a placebo indica que la intervención nunca ha sido comparada de manera directa con otros tratamientos activos. En cambio, si existen varias conexiones, el lector puede comprobar la solidez y la riqueza de la red de evidencia.

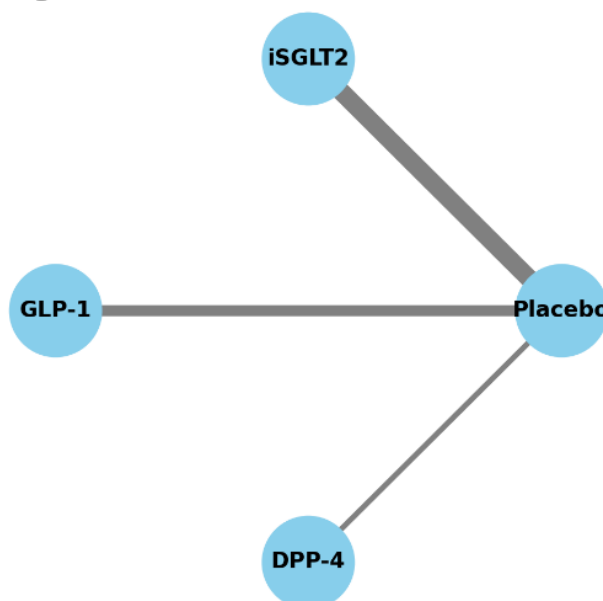
En términos prácticos, el gráfico de red permite responder rápidamente a tres preguntas críticas:

- ¿Qué intervenciones han sido comparadas directamente?
- ¿Cuáles solo se conectan a través de comparaciones indirectas?
- ¿Qué tan equilibrada es la red en cuanto a número y tamaño de los estudios?

Veamos un ejemplo aplicado:

Figura 4. Gráfico de red en un metaanálisis ficticio de tratamientos para diabetes tipo 2.

Ejemplo de gráfico de red en un metaanálisis (datos ficticios)



En este ensayo clínico de tres familias de fármacos para diabetes tipo 2: inhibidores de SGLT2, agonistas de GLP-1 e inhibidores de DPP-4. Existen ensayos que comparan cada grupo de fármacos con placebo, pero no todos los comparan entre sí. El gráfico de red muestra tres nodos (iSGLT2, GLP-1 y DPP-4) unidos a un nodo central (placebo). El grosor de las líneas variaría según cuántos estudios respalden cada comparación: una línea más gruesa entre iSGLT2 y placebo reflejaría que existen más ensayos en esa comparación que en la de DPP-4 frente a placebo. La fortaleza del metaanálisis en red radica precisamente en esta estructura: incluso si no hay un ensayo directo que compare iSGLT2 contra GLP-1, la red permite estimar indirectamente esa comparación gracias al “puente” común del placebo.

CONSISTENCIA

La consistencia es el pilar que sostiene la validez de un metaanálisis en red. Recordemos que la fuerza de este tipo de análisis radica en la posibilidad de combinar comparaciones directas e indirectas. Pero para que las conclusiones sean confiables, ambas deben llegar a resultados similares.

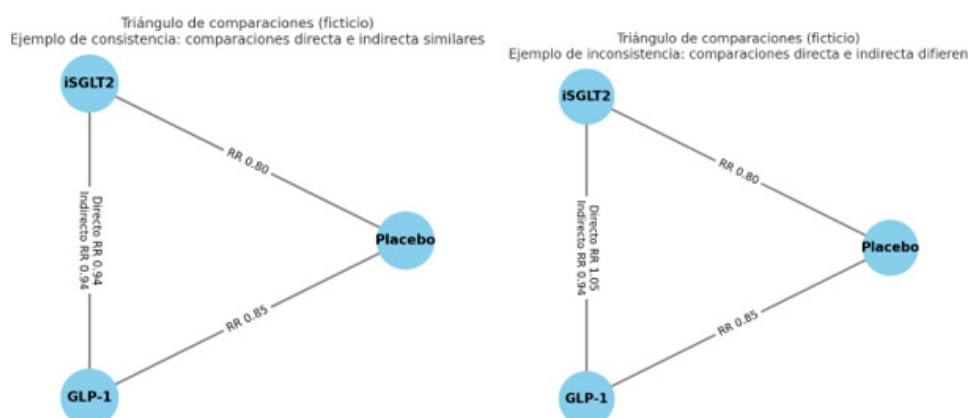
En términos simples, hay consistencia cuando la estimación de efecto obtenida a partir de los ensayos que comparan directamente dos tratamientos coincide con la que se obtiene de manera indirecta a través de un tercero. Si los resultados difieren, hablamos de inconsistencia o incoherencia, lo que puede comprometer seriamente la interpretación del metaanálisis.

Un ejemplo puede clarificarlo. Supongamos que los ensayos clínicos muestran que el inhibidor de SGLT2 reduce un 20 % los eventos cardiovasculares frente a placebo, y que los agonistas de GLP-1 reducen un 15 % esos mismos eventos frente a placebo. Por inferencia indirecta, cabría esperar que la comparación entre SGLT2 y GLP-1 muestre una diferencia cercana a ese 5 %. Si, sin embargo, los pocos estudios que comparan directamente ambos tratamientos encuentran que no hay diferencia o incluso que GLP-1 es superior, surge una inconsistencia.

Existen diferentes formas de evaluar esta consistencia. Algunas son locales, centradas en un lazo específico de la red (por ejemplo, un triángulo de comparaciones entre tres intervenciones), y otras son globales, que valoran la coherencia de toda la red. Entre las pruebas más usadas se encuentran el método de Bucher, que compara la diferencia entre la evidencia directa e indirecta en un triángulo cerrado, y los modelos de inconsistencia global, que analizan toda la red de manera simultánea.

Para el lector crítico, no es necesario dominar todos los métodos estadísticos, pero sí reconocer un punto esencial: si la red muestra inconsistencias importantes, la validez de sus conclusiones disminuye. En tales casos, el ranking de tratamientos puede ser engañoso y la certeza de la evidencia debería degradarse.

En la práctica, la consistencia se interpreta como una condición de credibilidad: cuando está presente, el NMA amplía el alcance de la evidencia; cuando falta, obliga a tomar las conclusiones con cautela y a priorizar los resultados de comparaciones directas.



Cada nodo representa una intervención (placebo, iSGLT2, GLP-1) y las líneas muestran las comparaciones disponibles. En el panel izquierdo, la estimación directa de iSGLT2 vs. GLP-1 (RR 0.94) coincide con la estimación indirecta a través del placebo (RR 0.94), lo que refleja consistencia en la red. En el panel derecho, la comparación directa (RR 1.05) difiere de la estimación indirecta (RR 0.94), lo que evidencia inconsistencia. La consistencia es un requisito fundamental para confiar en las conclusiones de un NMA; cuando falta, las comparaciones indirectas deben interpretarse con cautela y se debe dar mayor peso a la evidencia directa.

RANKINGS (SUCRA)

Uno de los aportes distintivos del metaanálisis en red es la posibilidad de ordenar los tratamientos según su probabilidad de ser el más eficaz o el más seguro. Esto se hace calculando, para cada intervención, la probabilidad acumulada de ocupar cada posición dentro del ranking.

La medida más utilizada para expresar este resultado es el SUCRA (Surface Under the Cumulative Ranking curve). Este indicador varía entre 0 y 1:

- Un valor cercano a 1 significa que la intervención tiene alta probabilidad de estar entre las mejores opciones.
- Un valor cercano a 0 indica que probablemente se ubique entre las menos eficaces o menos seguras.

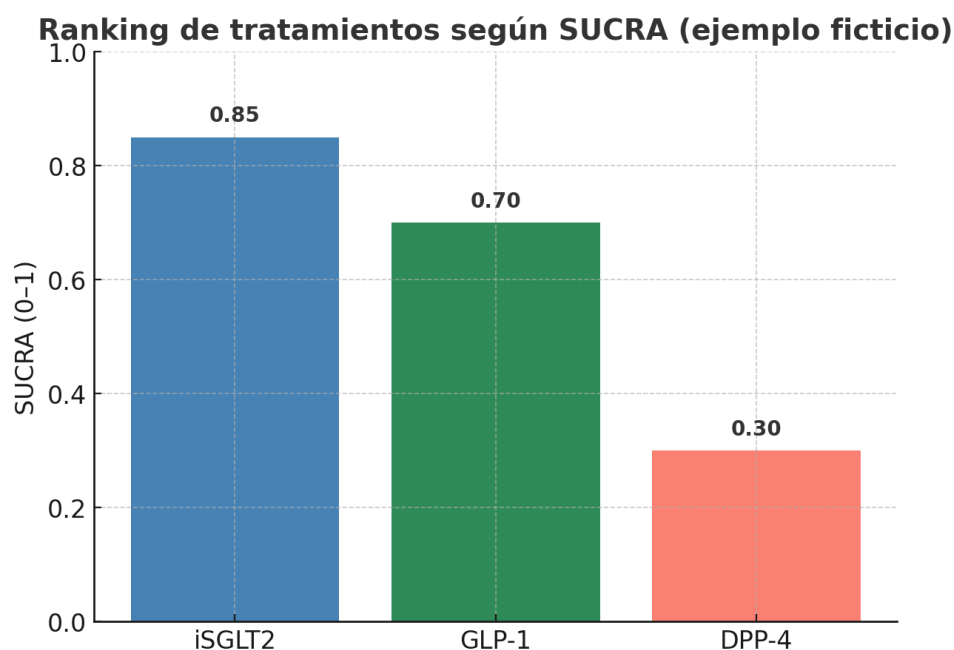
Por ejemplo, en un metaanálisis en red sobre antidiabéticos, los inhibidores de SGLT2 podrían obtener un SUCRA de 0.85 para la reducción de eventos cardiovasculares, los agonistas de GLP-1 un 0.70 y los inhibidores de DPP-4 un 0.30. Esto sugiere que iSGLT2 es la opción más sólida, seguida de GLP-1, mientras que los DPP-4 tienen menor probabilidad de beneficio.

Es importante subrayar que el ranking no equivale a una verdad absoluta. Un SUCRA alto no significa que el tratamiento sea siempre superior, sino que, dentro del conjunto de estudios analizados, tiene mayor probabilidad de ser el mejor. Además, las diferencias entre intervenciones con SUCRA cercanos pueden no ser clínicamente relevantes.

Para la lectura crítica, el ranking debe interpretarse siempre junto con la calidad de la evidencia y con la magnitud absoluta del efecto. Un tratamiento puede aparecer primero en la jerarquía, pero si la certeza de la evidencia es baja (por ejemplo, debido a inconsistencia o riesgo de sesgo), esa posición debe relativizarse.

En síntesis, los rankings del NMA ofrecen una visión comparativa muy útil para la práctica clínica y la elaboración de guías, pero el lector crítico debe tomarlos como una herramienta complementaria, no como el único criterio para la toma de decisiones.

Figura 6. Ranking de tratamientos según SUCRA en un metaanálisis en red (ejemplo ficticio) para eventos cardiovasculares.



En esta figura se muestran los resultados del ranking de eficacia de tres familias de fármacos para diabetes tipo 2 (iSGLT2, GLP-1 y DPP-4) en la reducción de eventos cardiovasculares mayores, expresados mediante el indicador SUCRA (Surface Under the Cumulative Ranking curve).

Los inhibidores de SGLT2 presentan un SUCRA de 0.85, lo que indica una alta probabilidad de ocupar las primeras posiciones en el ranking. Los agonistas de GLP-1 alcanzan un valor intermedio de 0.70, lo que los sitúa como una alternativa también favorable, aunque por debajo de iSGLT2. Finalmente, los inhibidores de DPP-4 muestran un SUCRA de 0.30, reflejando una menor probabilidad de ser considerados entre las opciones más eficaces.

LEAGUE TABLE

Una de las formas más claras de presentar los resultados de un NMA es la llamada league

table o tabla de comparaciones múltiples. Se trata de una matriz en la que cada fila y cada columna corresponden a un tratamiento, y en las celdas se presentan las comparaciones par a par.

Esta tabla permite que el lector vea de un solo vistazo todas las estimaciones posibles, incluso aquellas que provienen de comparaciones indirectas. Normalmente se reporta la medida de efecto (RR, OR, HR o diferencia de medias) junto con su intervalo de confianza al 95 %.

La interpretación es sencilla pero poderosa:

- Cada celda indica cómo se comporta el tratamiento de la fila en comparación con el de la columna.
- Un valor menor a 1 (en RR u OR) favorece al tratamiento de la fila; un valor mayor a 1 favorece al de la columna.
- La diagonal de la tabla queda vacía o marcada con “–” porque representa la comparación de un tratamiento consigo mismo.
- En artículos internacionales, estas tablas suelen ir acompañadas de colores o sombreados que ayudan a visualizar rápidamente qué tratamiento resulta superior.

Tabla 7. Comparaciones entre tratamientos en un metaanálisis en red (ejemplo ficticio) para eventos cardiovasculares.

	iSGLT2	GLP-1	DPP-4
iSGLT2	–	0.94 (0.88–0.99)	0.80 (0.72–0.89)
GLP-1	1.06 (1.01–1.14)	–	0.85 (0.77–0.94)
DPP-4	1.25 (1.12–1.39)	1.18 (1.06–1.30)	–

Fuente: Elaboración propia.

La tabla muestra de manera simultánea todas las comparaciones entre los tres tratamientos incluidos en el metaanálisis en red. En este caso ficticio, los inhibidores de SGLT2, los agonistas de GLP-1 y los inhibidores de DPP-4 fueron evaluados en la reducción de eventos cardiovasculares mayores, utilizando el riesgo relativo como medida de efecto.

Al analizar la fila correspondiente a los inhibidores de SGLT2, se observa que estos presentan un riesgo relativo de 0.94 frente a GLP-1, lo que indica una ligera ventaja a favor de los SGLT2, con un intervalo de confianza que descarta la ausencia de efecto. En la comparación contra los inhibidores de DPP-4, el riesgo relativo de 0.80 refleja un beneficio aún más marcado de los SGLT2, lo que refuerza su posición como la opción más eficaz dentro de la red.

Cuando se examinan los agonistas de GLP-1, se aprecia que, en comparación con los DPP-4, también ofrecen una reducción significativa del riesgo (RR 0.85), situándolos en una posición intermedia: menos eficaces que los SGLT2, pero claramente superiores a los DPP-4.

Por último, los inhibidores de DPP-4 aparecen sistemáticamente como la opción menos favorable, ya que frente a SGLT2 y GLP-1 presentan riesgos relativos superiores a 1, lo que implica una mayor probabilidad de eventos cardiovasculares en comparación con sus competidores.

En conjunto, la tabla permite visualizar de un vistazo la jerarquía terapéutica: los inhibidores de SGLT2 se perfilan como la intervención con mejores resultados, seguidos de los agonistas de GLP-1, mientras que los inhibidores de DPP-4 se ubican en el último lugar. Este tipo de presentación es especialmente útil para el lector clínico, pues ofrece un panorama completo de las comparaciones en un solo cuadro y facilita la interpretación crítica de la red de evidencia.

FOREST PLOT DE COMPARACIONES CLAVE

Aunque el metaanálisis en red genera un gran número de comparaciones posibles, no todas son igualmente relevantes para el lector clínico. Por ello, muchos artículos presentan forestplots resumidos con las comparaciones más importantes, generalmente:

- Cada tratamiento activo frente a placebo o control.
- Las comparaciones entre los tratamientos mejor posicionados en el ranking.

El formato es el mismo que en un metaanálisis clásico, la utilidad de este gráfico es doble. En primer lugar, permite visualizar rápidamente cuáles intervenciones ofrecen un beneficio clínico significativo frente al control. En segundo lugar, facilita la comparación directa entre los tratamientos que compiten por el primer lugar en el ranking, lo que ayuda a tomar decisiones prácticas en guías y protocolos clínicos.

6.4 SÍNTESIS NARRATIVAS

La síntesis narrativa es el método más utilizado cuando los estudios disponibles no pueden combinarse cuantitativamente en un metaanálisis. Esto suele ocurrir porque los ensayos son demasiado heterogéneos en cuanto a población, intervenciones, desenlaces o metodología, o porque la pregunta de investigación es demasiado amplia para permitir una síntesis estadística coherente.

En este enfoque, los resultados se presentan en forma de texto, a menudo acompañados de tablas que resumen las características de los estudios y sus hallazgos principales. El objetivo no es obtener un estimador global, sino describir las tendencias generales de la evidencia, identificar coincidencias y discrepancias, y señalar las limitaciones y vacíos de conocimiento.

Por ejemplo, una síntesis narrativa sobre terapias no farmacológicas para diabetes tipo 2 podría concluir que la mayoría de los estudios apoyan el beneficio de la dieta mediterránea y el ejercicio regular en el control glucémico, aunque con diferencias en la magnitud del efecto y en la calidad metodológica. En lugar de un único número combinado, el lector recibe un panorama contextualizado que le permite comprender mejor la diversidad de enfoques.

La ventaja de la síntesis narrativa es su flexibilidad: permite incluir estudios de distintos diseños y responder preguntas amplias. Sin embargo, su principal limitación es la subjetividad, ya que depende en gran medida de cómo los autores seleccionen, organicen e interpreten la información. Esto implica que, para tener credibilidad, debe estar guiada por un protocolo claro y transparente, con criterios explícitos de inclusión y exclusión.

Para el lector crítico, una buena síntesis narrativa debería responder a preguntas como:

- ¿Qué tratamientos o intervenciones muestran resultados consistentes en la mayoría de los estudios?
- ¿Dónde se encuentran las discrepancias y cuáles podrían ser sus causas?
- ¿Qué lagunas en la evidencia justifican nuevas investigaciones?

En conclusión, la síntesis narrativa es una herramienta valiosa cuando el metaanálisis no es posible, pero sus conclusiones deben interpretarse con mayor cautela, reconociendo el riesgo de sesgo y la ausencia de un análisis cuantitativo formal.

6.5 METAANÁLISIS DE PRUEBAS DIAGNOSTICAS

Los metaanálisis de pruebas diagnósticas tienen como objetivo sintetizar la exactitud de una prueba en diferentes contextos clínicos y poblaciones. A diferencia de los metaanálisis de intervenciones, en los que el desenlace suele ser un riesgo relativo o una diferencia de medias, aquí el interés se centra en medidas como la sensibilidad, la especificidad y las razones de verosimilitud (likelihood ratios), para más explicación de estas pruebas consultar el Capítulo 10 de Pruebas diagnósticas.

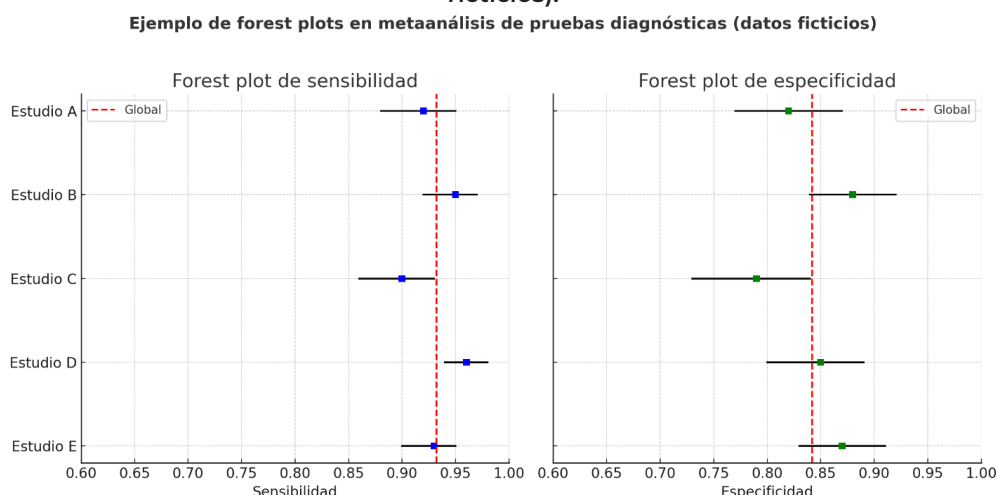
La síntesis no puede hacerse con métodos simples de promedios, ya que sensibilidad y especificidad están correlacionadas: cuando una aumenta, la otra suele disminuir. Por ello, se utilizan modelos más sofisticados, como el modelo bivariado de efectos aleatorios o las curvas ROC jerárquicas (HSROC). Estas herramientas permiten estimar el rendimiento global de la prueba considerando la variabilidad entre estudios.

FOREST PLOT DE SENSIBILIDAD Y ESPECIFICIDAD:

Los metaanálisis de pruebas diagnósticas suelen presentar dos forestplots paralelos: uno para la sensibilidad y otro para la especificidad. En cada gráfico, cada estudio aparece representado por un cuadrado (estimación puntual) acompañado de una línea horizontal (intervalo de confianza al 95 %). Al final del gráfico, se muestra un diamante que representa la estimación global combinada.

La interpretación es similar a la de cualquier forestplot, pero con una lógica centrada en el diagnóstico:

- En el forestplot de sensibilidad, los estudios se alinean alrededor de su capacidad para identificar correctamente a los pacientes con la enfermedad. Una estimación global alta (por ejemplo, >90 %) indica que la prueba es muy eficaz para detectar enfermos.
- En el forestplot de especificidad, los estudios se alinean en torno a su capacidad para identificar correctamente a quienes no tienen la enfermedad. Una estimación global elevada (por ejemplo, >85 %) significa que la prueba es útil para descartar sanos.

Figura 7. Forest plots de sensibilidad y especificidad en un metaanálisis de pruebas diagnósticas (datos ficticios).

En el gráfico de la izquierda se observa la sensibilidad de la troponina ultrasensible para el diagnóstico de infarto agudo de miocardio en distintos estudios. La mayoría de los ensayos se agrupan por encima del 90 %, con intervalos de confianza relativamente estrechos. La estimación global, representada por la línea roja discontinua, confirma que la prueba es altamente sensible y, por lo tanto, muy poco probable que pase por alto un caso verdadero de infarto.

El gráfico de la derecha muestra la especificidad, es decir, la capacidad de la prueba para identificar correctamente a los individuos sanos. En este caso se observa una mayor variabilidad entre los estudios, con valores que oscilan entre el 79 % y el 88 %. La estimación global se sitúa en torno al 85 %, lo que indica que, si bien la prueba es bastante específica, aún existe un número no despreciable de falsos positivos.

En conjunto, las figuras ilustran una conclusión clínica importante: la troponina ultrasensible es muy eficaz para descartar infarto cuando el resultado es negativo, pero su interpretación debe complementarse con la clínica y otras pruebas debido a la presencia de falsos positivos en algunos contextos.

CURVAS ROC JERÁRQUICAS (HSROC)

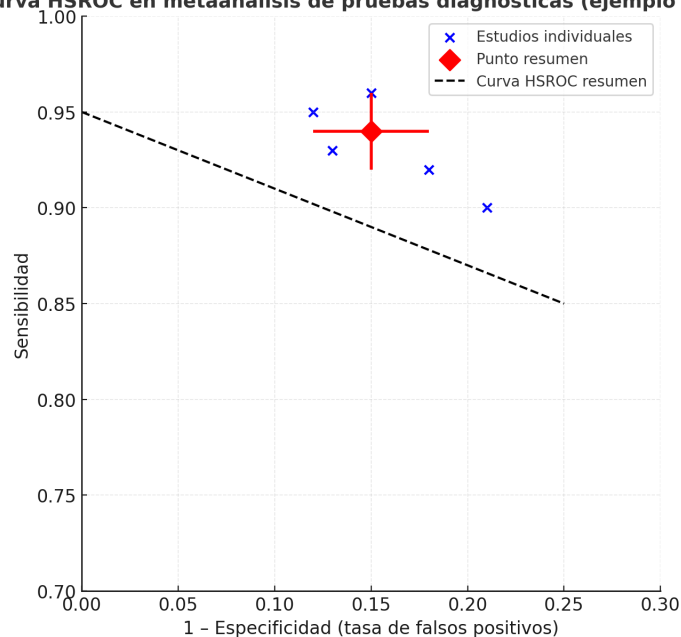
En los metaanálisis de pruebas diagnósticas, la presentación de forestplots de sensibilidad y especificidad resulta muy útil, pero tiene una limitación: obliga al lector a mirar dos gráficos por separado. Para superar este problema, se utilizan las curvas ROC jerárquicas (HSROC, por sus siglas en inglés), que permiten visualizar la exactitud diagnóstica de una prueba en un solo gráfico bidimensional.

En este tipo de representación, cada estudio individual se ubica en un plano cartesiano donde el eje X corresponde a $1 - \text{especificidad}$ (tasa de falsos positivos) y el eje Y a la sensibilidad. De esta manera, los estudios con mejor rendimiento aparecen en la parte superior izquierda de la gráfica (alta sensibilidad y alta especificidad).

La gran ventaja del modelo HSROC es que ajusta estadísticamente la correlación entre sensibilidad y especificidad, permitiendo trazar una curva resumen que refleja el rendimiento global de la prueba. Además, se calcula un punto resumen con su intervalo de confianza, lo que facilita cuantificar la exactitud media.

La interpretación crítica de una curva HSROC debe responder a tres preguntas:

- ¿Dónde se ubica el punto resumen? → cuanto más cercano al ángulo superior izquierdo, mejor es la prueba.
- ¿Qué tan anchos son los intervalos de confianza? → intervalos estrechos sugieren homogeneidad y mayor certeza; intervalos amplios reflejan variabilidad entre estudios.
- ¿Cómo se distribuyen los puntos de los estudios? → una nube de puntos muy dispersa indica heterogeneidad, lo que puede limitar la aplicabilidad de la prueba en distintos contextos clínicos.

Curva HSROC en metaanálisis de pruebas diagnósticas (ejemplo ficticio)

Se observa que la mayoría de los estudios se concentran en la parte superior izquierda de la gráfica, lo que refleja una alta sensibilidad ($\geq 90\%$) y una especificidad moderada (80–90 %). El rombo rojo señala el punto resumen combinado (sensibilidad aproximada del 94 % y especificidad del 85 %), acompañado de sus intervalos de confianza. Este punto cercano al ángulo superior izquierdo indica que, en conjunto, la prueba tiene un rendimiento diagnóstico global muy favorable.

La curva discontinua muestra la trayectoria HSROC que resume el comportamiento esperado de la prueba a lo largo de distintos umbrales de decisión. La proximidad de esta curva al ángulo superior izquierdo es un indicador visual de la solidez del desempeño de la prueba.

Desde una perspectiva clínica, esta representación permite concluir que la troponina ultrasensible es una herramienta muy útil para descartar infarto en pacientes con dolor torácico, ya que rara vez omite un caso verdadero. Sin embargo, la dispersión de los puntos entre estudios recuerda que existe cierta heterogeneidad, y que la especificidad no es perfecta: en algunos contextos clínicos la prueba puede generar falsos positivos que deben interpretarse con cautela junto a la historia clínica y otros exámenes complementarios.

6.5 MÍNIMA DIFERENCIA CLÍNICAMENTE IMPORTANTE

En la interpretación de un metaanálisis no basta con comprobar si los resultados son estadísticamente significativos. El verdadero desafío para el lector crítico es preguntarse si esos resultados son clínicamente relevantes. Aquí es donde entra en juego la MCID (Minimal Clinically Important Difference), o mínima diferencia clínicamente importante.

La MCID se define como la menor magnitud de cambio en un desenlace que los pacientes perciben como beneficiosa o que justifica una decisión clínica. Dicho de otra forma, es el umbral que separa un efecto “estadísticamente significativo” de un efecto que realmente tiene sentido en la práctica.

Un ejemplo ayuda a clarificarlo. Supongamos que un metaanálisis de un nuevo analgésico muestra una reducción media del dolor de 0.5 puntos en una escala de 0 a 10, con un intervalo de confianza estrecho y un valor $p < 0.01$. Desde el punto de vista estadístico, el resultado es sólido. Pero si la MCID definida para el dolor en esa escala es de 1.5 puntos, la conclusión clínica es distinta: el fármaco sí reduce el dolor, pero el cambio es tan pequeño que probablemente los pacientes no lo perciban como relevante.

En los metaanálisis, la MCID se utiliza de varias formas. Puede guiar la interpretación de diferencias de medias (MD o SMD), indicando si el efecto supera o no el umbral clínicamente significativo. También se emplea para valorar medidas de riesgo: por ejemplo, una reducción absoluta del riesgo de 2 % en eventos cardiovasculares puede ser estadísticamente significativa, pero si el umbral clínicamente relevante se considera en torno al 5 %, el resultado debe interpretarse

con cautela.

La MCID no es un valor universal, sino que depende del contexto clínico, la población y el desenlace. En muchos casos proviene de estudios de validación, consenso de expertos o encuestas a pacientes. Por ello, su correcta interpretación exige que el lector sepa qué MCID se ha definido para cada desenlace y si el metaanálisis lo supera.

En conclusión, la MCID ayuda a distinguir entre lo estadísticamente significativo y lo clínicamente relevante. Un metaanálisis bien interpretado debe siempre considerar este umbral, de manera que los resultados no solo sean confiables en términos numéricos, sino también útiles para orientar decisiones clínicas reales.

6.6 SISTEMA GRADE

El sistema GRADE (Grading of Recommendations Assessment, Development and Evaluation) se ha convertido en el estándar internacional para valorar la calidad o certeza de la evidencia científica. Su importancia radica en que no se limita a describir si un resultado es “estadísticamente significativo”, sino que se centra en la confianza que podemos tener en ese resultado y en su relevancia para la práctica clínica.

En términos sencillos, GRADE responde a la pregunta: ¿qué tan seguros estamos de que el efecto estimado por un metaanálisis refleja la realidad clínica?

Niveles de certeza: GRADE clasifica la evidencia en cuatro niveles:

- Alta: tenemos gran seguridad en que el efecto verdadero está muy próximo al estimado. En la práctica, significa que nuevos estudios es muy poco probable que cambien la conclusión.
- Moderada: existe confianza razonable en la estimación, pero persiste la posibilidad de que estudios futuros modifiquen la magnitud o incluso la dirección del efecto.
- Baja: el efecto real puede ser sustancialmente diferente al estimado, y la conclusión debe tomarse con cautela.
- Muy baja: la incertidumbre es tan grande que el efecto estimado puede no reflejar en absoluto la realidad.

Esta clasificación se realiza por cada desenlace crítico y no de manera global para todo el metaanálisis. Así, un mismo estudio puede tener certeza alta para mortalidad y certeza baja para calidad de vida.

Existen factores que disminuyen la certeza, la calificación de certeza comienza en un nivel distinto según el tipo de estudio:

- Los ensayos clínicos aleatorizados parten en “alta certeza”.
- Los estudios observacionales parten en “baja certeza”.

A partir de esa base, la certeza puede descender por cinco motivos:

- Riesgo de sesgo: estudios con ocultamiento inadecuado, pérdidas significativas, falta de cegamiento o análisis selectivo de resultados.
 - Inconsistencia: resultados heterogéneos entre estudios, sin explicación clara.
 - Indirectitud: la población, intervención, comparador o desenlace no corresponden exactamente a la pregunta clínica. Por ejemplo, usar estudios en adultos jóvenes para responder una pregunta sobre adultos mayores.
 - Imprecisión: intervalos de confianza amplios, escaso número de eventos o tamaños muestrales reducidos.
 - Sesgo de publicación: sospecha de que solo se han publicado estudios con resultados positivos.
- Así mismo existen factores que pueden aumentar la certeza, en el caso de estudios observacionales, la certeza puede incrementarse si se cumplen ciertos criterios:

- Un efecto muy grande (ejemplo: reducción de mortalidad del 80 %).
- Evidencia de relación dosis-respuesta.
- Presencia de factores de confusión que, de estar presentes, reforzarían aún más el efecto observado.

El valor añadido de GRADE: Lo que distingue a GRADE de otros sistemas de clasificación es que no se centra solo en los aspectos estadísticos o metodológicos, sino en el valor clínico de la evidencia. Esto significa que no basta con saber si un tratamiento funciona; debemos conocer también la confianza que podemos tener en ese resultado para aplicarlo a nuestros pacientes.

Por ejemplo, un metaanálisis puede mostrar que un fármaco reduce la mortalidad con un RR de 0.80. Pero si la certeza es baja, el lector crítico sabrá que esa reducción debe interpretarse con prudencia, pues estudios futuros podrían modificarla. En cambio, si la certeza es alta, la conclusión gana peso para sustentar recomendaciones firmes en una guía clínica.

TABLAS SUMMARY OF FINDINGS (SOF)

La aplicación práctica de GRADE se traduce en las tablas Summary of Findings, que resumen los desenlaces más importantes con su estimación puntual, su intervalo de confianza y el nivel de certeza. Estas tablas, empleadas habitualmente en revisiones Cochrane, ofrecen en una sola página la información más relevante para la toma de decisiones.

Su valor añadido es que presentan tanto medidas relativas (como el RR u OR) como efectos absolutos, expresados en reducción o aumento de eventos por cada 1,000 pacientes tratados. Esto acerca los resultados al terreno clínico: no es lo mismo saber que un fármaco reduce el riesgo un 20 % (medida relativa), que entender que esto implica 20 eventos menos por cada 1,000 pacientes (medida absoluta).

En síntesis: GRADE transforma la lectura de un metaanálisis en un ejercicio más clínico que estadístico. Invita al lector a preguntarse no solo si el tratamiento funciona, sino qué tan seguro puedo estar de ese resultado y qué significa en términos de beneficio real para mis pacientes.

Como ejemplo se puede mencionar:

Tabla 8. Summary of Findings (SoF): inhibidores de SGLT2 frente a placebo en eventos cardiovasculares mayores (ejemplo ficticio)

Desenlace crítico	Riesgo con placebo (por 1,000)	Riesgo con iSGLT2 (por 1,000)	Efecto relativo (RR, IC95%)	Reducción absoluta (por 1,000)	Certeza de la evidencia (GRADE)
Eventos cardiovasculares mayores (MACE: muerte CV, IAM, ACV)	100	80	0.80 (0.72–0.89)	–20 (–28 a –11)	Alta
Hospitalización por insuficiencia cardíaca	80	60	0.75 (0.66–0.86)	–20 (–27 a –11)	Moderada (inconsistencia leve)
Mortalidad por todas las causas	120	100	0.83 (0.74–0.94)	–20 (–31 a –7)	Moderada (imprecisión)
Eventos adversos graves	150	155	1.03 (0.95–1.12)	+5 (–8 a +18)	Alta

Fuente: Elaboración propia.

Para su interpretación: La tabla 8 resume los desenlaces clínicamente más relevantes del metaanálisis ficticio. Se observa que los iSGLT2 reducen de manera significativa los eventos cardiovasculares mayores, con una certeza alta, lo que otorga gran confianza al resultado. La hospitalización por insuficiencia cardíaca y la mortalidad por todas las causas también se reducen, aunque la certeza es moderada, debido a cierta heterogeneidad y a intervalos de confianza más amplios.

En cuanto a la seguridad, los eventos adversos graves no mostraron diferencias significativas entre iSGLT2 y placebo, con certeza alta, lo que aporta tranquilidad al considerar el balance riesgo-beneficio.

En conjunto, esta tabla ejemplifica la utilidad de los formatos Summary of Findings: condensan la evidencia cuantitativa y la calidad de la misma en un formato claro, permitiendo al clínico valorar no solo si un fármaco funciona, sino también qué tan sólida es la confianza en ese resultado y cuán aplicable es en la práctica.

6.7 ERRORES COMUNES DE INTERPRETACIÓN

La lectura crítica de revisiones sistemáticas y metaanálisis exige no solo conocer la metodología, sino también reconocer las trampas interpretativas más frecuentes. Identificarlas ayuda al lector a evitar conclusiones erróneas y a valorar con justicia la solidez de la evidencia.

Confundir significancia estadística con relevancia clínica: Un error habitual es considerar que un resultado estadísticamente significativo es automáticamente clínicamente importante. Por ejemplo, un metaanálisis puede mostrar que un fármaco reduce la hemoglobina glucosilada en 0.3 %, con un intervalo de confianza estrecho y un valor $p < 0.001$. Sin embargo, si la mínima diferencia clínicamente importante (MCID) para este desenlace es de 0.5 %, el beneficio existe en números, pero difícilmente sea percibido por el paciente o cambie su pronóstico.

Ignorar la heterogeneidad: La presencia de heterogeneidad elevada ($I^2 > 50 \%$) debería despertar cautela, pues indica que los estudios no son completamente comparables. Un error frecuente es interpretar el estimador combinado como una “verdad universal”, sin reparar en que los resultados pueden variar entre contextos. En estos casos, la síntesis narrativa, los subgrupos o

la metarregresión cobran más importancia que el valor global.

Tomar los rankings de NMA como verdades absolutas: En los metaanálisis en red, los rankings de eficacia (ej. SUCRA) resultan muy atractivos porque ordenan tratamientos en forma de lista. Sin embargo, un error común es asumir que el primero en el ranking es siempre la mejor opción clínica. En realidad, la diferencia entre los dos primeros puede ser mínima o clínicamente irrelevante, y la certeza de la evidencia puede ser baja. Los rankings deben interpretarse como una guía orientativa, no como un resultado definitivo.

Pasar por alto el riesgo de sesgo y la calidad de los estudios: Un metaanálisis puede incluir decenas de ensayos, pero si la mayoría tiene alto riesgo de sesgo, el resultado combinado puede ser poco confiable. Es un error frecuente fijarse únicamente en el número de estudios sin atender a su calidad. El sistema GRADE y las herramientas como RoB 2 ayudan a contextualizar este problema.

Sobrevalorar la estabilidad de los resultados: Otro error común es asumir que un metaanálisis bien realizado siempre ofrece conclusiones sólidas. Sin embargo, el análisis de sensibilidad puede revelar que la exclusión de un solo estudio cambia radicalmente los resultados. Ignorar este aspecto puede llevar a una confianza excesiva en hallazgos frágiles.

Olvidar la aplicabilidad clínica: No todo resultado de un metaanálisis es automáticamente aplicable al paciente que tenemos enfrente. La indirectitud (diferencias entre la población estudiada y la nuestra) puede limitar la utilidad de la evidencia. Por ejemplo, un tratamiento probado en hospitales de alta complejidad puede no mostrar el mismo efecto en centros rurales con recursos limitados.

En síntesis: los metaanálisis son herramientas poderosas, pero no infalibles. El lector crítico debe recordar que significancia no equivale a relevancia clínica, que la heterogeneidad puede invalidar un estimador global, que los rankings son orientativos y que la aplicabilidad depende siempre del contexto.

7. SESGOS Y LIMITACIONES

Es fundamental recordar que no son inmunes a los problemas metodológicos ni a los sesgos. De hecho, su interpretación crítica requiere que el lector tenga presentes las limitaciones que pueden afectar tanto a la validez interna como a la aplicabilidad de sus resultados.

Uno de los problemas más importantes es el **sesgo de publicación**. Se sabe que los estudios con resultados positivos o estadísticamente significativos tienen mayor probabilidad de ser publicados en revistas de alto impacto, mientras que los estudios con resultados negativos o neutros tienden a permanecer inéditos o en repositorios poco accesibles. Esto genera un desequilibrio que favorece artificialmente los efectos beneficiosos de una intervención. Herramientas como el funnelplot o la prueba de Egger que vimos en acápites previos permiten detectar la presencia de este sesgo, y métodos de ajuste como el trim and fill ofrecen una aproximación de lo que sucedería si los estudios faltantes se incorporaran.

Otro aspecto crítico es el **sesgo de selección**, que puede ocurrir si la estrategia de búsqueda bibliográfica es limitada. Por ejemplo, una revisión que solo utiliza PubMed y restringe la búsqueda al idioma inglés corre el riesgo de dejar fuera estudios relevantes publicados en otras bases de datos o en idiomas distintos. La omisión de literatura gris (tesis, informes técnicos, actas de congresos) puede también contribuir a este sesgo, reduciendo la exhaustividad del análisis. De ahí la importancia de registrar un protocolo previo, donde la estrategia de búsqueda quede claramente definida.

La **heterogeneidad** clínica y metodológica representa otra limitación frecuente. En muchas ocasiones, los estudios incluidos difieren en características de los pacientes, dosis de la intervención, comparadores o definiciones de desenlaces. Cuando estas diferencias son notables, el estimador global puede resultar engañoso. Un ejemplo clásico se observa en los metaanálisis de prevención cardiovascular, donde mezclar poblaciones de bajo riesgo con pacientes de alto riesgo puede diluir el efecto real y dar lugar a conclusiones erróneas.

También es necesario considerar el **sesgo de reporte selectivo**. Algunos estudios primarios informan solo aquellos desenlaces que muestran un beneficio, omitiendo resultados desfavorables o neutros. Si los revisores no contrastan los artículos con los protocolos originales o los registros de ensayos clínicos, este sesgo pasa desapercibido y contamina el metaanálisis.

Más allá de los sesgos, existen **limitaciones metodológicas** que deben tenerse en cuenta. Un metaanálisis no puede superar la calidad de los estudios que lo integran: si estos son pequeños, con alto riesgo de sesgo o metodológicamente débiles, la estimación global será igualmente

frágil. Del mismo modo, la imprecisión estadística, reflejada en intervalos de confianza amplios o en un bajo número de eventos, disminuye la certeza de las conclusiones. Incluso las decisiones estadísticas, como optar entre un modelo de efectos fijos o uno de efectos aleatorios, pueden cambiar de forma importante los resultados.

Finalmente, es indispensable reflexionar sobre las **limitaciones prácticas**. Un metaanálisis puede mostrar un beneficio convincente en un escenario ideal, pero ser de poca aplicabilidad en el contexto real del lector. Por ejemplo, un metaanálisis que evalúa procedimientos quirúrgicos realizados en hospitales de excelencia de países desarrollados podría no reflejar la misma magnitud de beneficio en hospitales de segundo nivel de América Latina, donde las condiciones de infraestructura, entrenamiento del personal y disponibilidad de recursos son diferentes.

En resumen, los sesgos y limitaciones de las revisiones sistemáticas y metaanálisis nos recuerdan que, aunque se trate de estudios de máximo nivel de evidencia, no deben interpretarse como verdades absolutas. El lector crítico debe preguntarse siempre: ¿existen estudios que no fueron publicados? ¿la búsqueda bibliográfica fue exhaustiva? ¿los estudios incluidos son comparables entre sí? ¿la calidad metodológica es adecuada? y, sobre todo, ¿estos resultados son aplicables a mis pacientes? Solo con este ejercicio reflexivo se puede aprovechar todo el valor de la síntesis de evidencia sin caer en interpretaciones simplistas o engañosas.

8. COMPARACIÓN CON OTROS ESTUDIOS

Un aspecto esencial en la lectura crítica de revisiones sistemáticas y metaanálisis es la comparación con trabajos previos que hayan abordado la misma pregunta clínica o temas relacionados. Este ejercicio permite al lector ubicar la nueva evidencia en el contexto del conocimiento acumulado, identificar consistencias o discrepancias y comprender mejor las razones detrás de los hallazgos.

En algunos campos, puede haber múltiples revisiones sistemáticas publicadas con pocos años de diferencia. Aunque todas partan de una misma pregunta clínica, los resultados pueden variar por diferencias metodológicas: la amplitud de la búsqueda bibliográfica, los criterios de inclusión, el manejo de la heterogeneidad, el tipo de análisis estadístico o el enfoque en determinados subgrupos. Un lector crítico debe aprender a detectar estos matices, porque de ellos depende la interpretación final.

Por ejemplo, en el caso de los inhibidores de SGLT2 en pacientes con diabetes tipo 2, una revisión publicada en 2018 mostró una reducción significativa de la hospitalización por insuficiencia cardíaca, pero no de la mortalidad por todas las causas. Años más tarde, con la incorporación de nuevos ensayos clínicos, una revisión de 2021 encontró también un beneficio claro en mortalidad, lo que cambió la percepción del impacto clínico de estos fármacos. Este contraste ilustra cómo el tiempo y la acumulación de estudios modifican las conclusiones de los metaanálisis.

También es posible encontrar discrepancias entre revisiones que utilizan metodologías diferentes. Una revisión Cochrane, por su naturaleza estricta, puede excluir estudios de menor calidad y reportar un efecto más conservador. En cambio, una revisión narrativa o un metaanálisis con criterios más flexibles podría incluir mayor cantidad de estudios y sugerir un efecto más pronunciado, pero con menor certeza.

La comparación no debe limitarse a la magnitud del efecto, sino también a la certeza de la evidencia y a la aplicabilidad clínica. Revisiones realizadas en contextos de alta especialización pueden mostrar resultados distintos a los observados en entornos de atención primaria. Asimismo, revisiones centradas en desenlaces duros (mortalidad, eventos cardiovasculares) pueden diferir de aquellas que priorizan desenlaces intermedios o de calidad de vida.

En síntesis, comparar una revisión sistemática o metaanálisis con otros estudios previos no es un ejercicio anecdótico, sino una parte fundamental de la lectura crítica. Permite al lector valorar si la nueva evidencia confirma lo ya conocido, si corrige limitaciones de estudios anteriores o si plantea preguntas aún sin resolver.

9. CHECKLIST

A lo largo de los últimos veinte años se han desarrollado múltiples herramientas y declaraciones que establecen estándares para el reporte, la calidad metodológica y la evaluación de la certeza de la evidencia. Este checklist resume las principales guías y recursos actualizados (PRISMA, AMSTAR, GRADE, entre otros) y está diseñado como una hoja de ruta práctica para el

lector. No reemplaza la lectura completa de cada documento, pero ofrece un marco confiable para identificar fortalezas y debilidades en cualquier revisión publicada.

Tabla 9. Principales guías y herramientas internacionales para lectura crítica de revisiones sistemáticas y metaanálisis

Área	Guía / Recurso	Año (últ. act.)	Aplicación práctica en la lectura crítica
Registro y protocolos	PROSPERO	2011–2025	Confirma si la revisión fue registrada de forma prospectiva para evitar sesgo selectivo y duplicación.
	PRISMA-P	2015	Evalúa si el protocolo describe claramente la pregunta, criterios de inclusión y plan de análisis.
	OSF Registries	–	Alternativa abierta de registro cuando PROSPERO no aplica.
Reporte principal	PRISMA 2020	2021	Comprueba transparencia en métodos, selección de estudios, síntesis y reporte de resultados (27 ítems).
	PRISMA-S	2021	Revisa si la estrategia de búsqueda bibliográfica es reproducible y completa.
	SWiM	2020	Evalúa la calidad de la síntesis narrativa cuando no es posible realizar metaanálisis.
Extensiones PRISMA	PRISMA-DTA	2018–2021	Revisa RS de pruebas diagnósticas: exactitud, HSROC, modelos bivariados.
	PRISMA-NMA	2015	Evalúa la transparencia en metaanálisis en red y supuestos de consistencia.
	PRISMA-ScR	2018	Valora revisiones de alcance (scoping reviews), justificación y cobertura.
	PRISMA-IPD	2015	Para metaanálisis de datos individuales, garantiza reporte sobre obtención y armonización de datos.
	MOOSE	2000	Complementa PRISMA en revisiones de estudios observacionales.
Riesgo de sesgo en estudios primarios	RoB 2	2019–2021	Juzga el riesgo de sesgo en ECA, dominio por dominio y desenlace por desenlace.
	ROBINS-I	2016	Evalúa sesgo en estudios no aleatorizados, con especial énfasis en confusión y comparabilidad.
	QUADAS-2	2011	Revisa estudios de pruebas diagnósticas (pacientes, prueba índice, referencia, tiempo).
	QUADAS-C	2022	Nueva herramienta para estudios que comparan directamente pruebas diagnósticas.
Calidad de la revisión	AMSTAR 2	2017	Valora la calidad metodológica de la revisión completa (16 dominios).
	ROBIS	2016	Examina el riesgo de sesgo de la propia revisión sistemática.
Certeza y aplicabilidad	GRADE + SoF	2023–2024	Clasifica la certeza de la evidencia (alta–muy baja) y resume resultados en tablas de desenlaces.
	CINeMA	2020	Estima confianza en comparaciones de metaanálisis en red (6 dominios).

Fuente: Elaboración propia.

10. UTILIDAD

Su utilidad práctica trasciende el ámbito académico y repercute directamente en la atención clínica, en la elaboración de guías de práctica y en la toma de decisiones en salud pública.

En la **práctica clínica individual**, estos estudios ayudan al médico a responder preguntas frecuentes: ¿cuál es el mejor tratamiento para reducir eventos cardiovasculares en pacientes con diabetes? ¿qué prueba diagnóstica ofrece mayor exactitud para detectar una enfermedad en fase temprana? En lugar de basarse en la experiencia personal o en un solo ensayo, el profesional cuenta con un panorama amplio que reúne múltiples estudios, ponderando su calidad y consistencia.

En el terreno de la **salud pública**, los metaanálisis son una base sólida para la planificación de programas sanitarios, la evaluación de nuevas tecnologías y la asignación de recursos. Al condensar grandes cantidades de datos, permiten estimar el impacto poblacional de una intervención y su relación costo-beneficio.

En **investigación**, las revisiones sistemáticas cumplen otra función valiosa: identificar lagunas en el conocimiento. Señalan qué preguntas han sido respondidas con suficiente certeza

y cuáles requieren estudios adicionales. Por ejemplo, una revisión puede mostrar que existe evidencia robusta para adultos jóvenes, pero insuficiente en adultos mayores o en poblaciones de bajos ingresos, lo que abre nuevas líneas de investigación.

Finalmente, su utilidad se refleja en la **práctica normativa**. Organismos internacionales como la OMS, el NICE en el Reino Unido y el USPSTF en Estados Unidos utilizan revisiones sistemáticas y metaanálisis como insumo principal para la elaboración de recomendaciones clínicas. Esto se debe a que proporcionan un marco transparente, reproducible y basado en evidencia para orientar políticas de salud y guías clínicas.

Mensaje clave para el lector: Las revisiones sistemáticas y los metaanálisis son más que un ejercicio académico; son una herramienta que conecta la investigación con la práctica clínica y las políticas sanitarias. Su verdadera utilidad radica en que permiten tomar decisiones mejor fundamentadas, más seguras para los pacientes y más eficientes para los sistemas de salud.

11. PREGUNTAS DE AUTOEVALUACIÓN

PREGUNTAS DE SELECCIÓN MÚLTIPLE

1. ¿Cuál es la principal ventaja de un metaanálisis frente a un ensayo clínico individual?
 - a) Permite generar hipótesis nuevas
 - b) Aumenta la potencia estadística al combinar estudios
 - c) Elimina por completo el sesgo de publicación
 - d) Evita la necesidad de evaluación crítica
2. El funnelplot se utiliza principalmente para:
 - a) Detectar heterogeneidad clínica
 - b) Identificar sesgo de publicación
 - c) Estimar el efecto combinado
 - d) Evaluar el riesgo de sesgo de cada estudio
3. En un forestplot de metaanálisis, el diamante al final del gráfico representa:
 - a) El intervalo de confianza más amplio
 - b) La heterogeneidad entre los estudios
 - c) La estimación global combinada
 - d) El estudio de mayor peso
4. Según el sistema GRADE, ¿qué significa que la certeza de la evidencia sea “baja”?
 - a) Que los resultados son irrelevantes clínicamente
 - b) Que es poco probable que futuros estudios cambien las conclusiones
 - c) Que el efecto real puede diferir sustancialmente del estimado
 - d) Que la revisión fue mal realizada
5. En el contexto de revisiones sistemáticas de pruebas diagnósticas, un valor de $LR- < 0.1$ indica:
 - a) Que la prueba es inútil para descartar enfermedad
 - b) Que la prueba es muy poderosa para descartar enfermedad
 - c) Que la prueba confirma fuertemente la enfermedad
 - d) Que existe heterogeneidad significativa

PREGUNTAS ABIERTAS

6. Explique por qué la significancia estadística en un metaanálisis no siempre implica relevancia clínica.
7. ¿Qué diferencias existen entre un metaanálisis en red (NMA) y un metaanálisis convencional?
8. Mencione tres estrategias de análisis de sensibilidad en metaanálisis y explique su utilidad.
9. ¿Qué aporta la tabla Summary of Findings (SoF) al lector clínico que no ofrecen los forestplots por sí solos?
10. Describa brevemente cómo la indirectitud puede limitar la aplicabilidad de un metaanálisis.

RESPUESTAS

1. b). El metaanálisis integra los resultados de múltiples estudios, lo que incrementa el tamaño muestral efectivo y mejora la precisión de las estimaciones. Esto permite detectar efectos que podrían no ser evidentes en estudios individuales.

2. b). El funnelplot evalúa la distribución de los estudios según su tamaño y efecto. La asimetría sugiere la ausencia de estudios, generalmente negativos, lo que indica posible sesgo de publicación.

3. c). El diamante representa el efecto global del metaanálisis; su centro indica la estimación puntual y su amplitud el intervalo de confianza combinado.

4. c). Una certeza baja en GRADE implica que existe incertidumbre importante y que futuros estudios pueden modificar significativamente el resultado.

5. b). Un $LR- < 0.1$ reduce considerablemente la probabilidad de enfermedad tras un resultado negativo, lo que hace a la prueba útil para descartar diagnóstico.

6. La significancia estadística indica que un resultado es poco probable que se deba al azar, pero no necesariamente implica relevancia clínica. Un efecto puede ser estadísticamente significativo pero de magnitud pequeña, sin alcanzar la mínima diferencia clínicamente importante (MCID), por lo que no genera un beneficio real para el paciente.

7. El metaanálisis convencional compara dos intervenciones de forma directa, mientras que el metaanálisis en red permite comparar múltiples intervenciones simultáneamente, integrando evidencia directa e indirecta. Esto permite estimar comparaciones no evaluadas directamente y generar rankings de tratamientos.

8. La exclusión de estudios con alto riesgo de sesgo permite evaluar si los resultados dependen de estudios de baja calidad. El análisis leave-one-out identifica si algún estudio influye de forma desproporcionada en el resultado global. La comparación entre modelos de efectos fijos y aleatorios permite evaluar si las conclusiones dependen del modelo estadístico utilizado. Estas estrategias ayudan a determinar la robustez de los resultados.

9. La tabla Summary of Findings integra el tamaño del efecto, los intervalos de confianza y la certeza de la evidencia (GRADE), además de presentar resultados en términos absolutos. Esto facilita la interpretación clínica, a diferencia del forestplot, que se centra únicamente en la estimación estadística.

10. La indirectitud ocurre cuando los estudios incluidos no representan adecuadamente la población, intervención o desenlace de interés. Esto limita la aplicabilidad de los resultados, ya que las conclusiones pueden no ser extrapolables al contexto clínico real del paciente.

REFERENCIAS

- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. Available at: <https://doi.org/10.1136/bmj.n71>.
- Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane handbook for systematic reviews of interventions*. Version 6.5 [updated Aug 2024]. Cochrane; 2024. Available at: <https://www.cochrane.org/handbook>.
- Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *J Med Libr Assoc*. 2021;109(2):174–200. Available at: <https://doi.org/10.5195/jmla.2021.962>.
- Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*. 2020;368:l6890. Available at: <https://doi.org/10.1136/bmj.l6890>.
- Nikolakopoulou A, Higgins JPT, Papakonstantinou T, Chaimani A, Giovane CD, Egger M, et al. CINeMA: An approach for assessing confidence in the results of a network meta-analysis. *PLoS Med*. 2020;17(4):e1003082. Available at: <https://doi.org/10.1371/journal.pmed.1003082>.
- McInnes MDF, Moher D, Thoms BD, McGrath TA, Bossuyt PM, Clifford T, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): The PRISMA-DTA statement. *JAMA*. 2018;319(4):388–96. Available at: <https://doi.org/10.1001/jama.2017.19163>.
- Shea BJ, Reeves BC, 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. *BMJ*. 2017;358:j4008. Available at: <https://doi.org/10.1136/bmj.j4008>.
- Whiting P, Savović J, Higgins JPT, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016;69:225–34. Available at: <https://doi.org/10.1016/j.jclinepi.2015.06.005>.
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177–88. Available at: [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2).
- Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629–34. Available at: <https://doi.org/10.1136/bmj.315.7109.629>.