

Revisiones Sistemáticas y Metanálisis de estudios epidemiológicos:

una guía para estudiantes de ciencias de la salud



BÚSQUEDA
SISTEMÁTICA



SELECCIÓN
RIGUROSA



ANÁLISIS
INTEGRADO



EVIDENCIA PARA
DECISIONES

**LUIS FELIPE HIGUITA GUTIÉRREZ
JAIBERTH ANTONIO CARDONA ARIAS**

“Revisiones Sistemáticas y Metanálisis de estudios epidemiológicos: una guía para estudiantes de ciencias de la salud” / Luis Felipe Higueta-Gutiérrez y Jaiberth Antonio Cardona-Arias. Santa Rosa de Osos: Editorial Católica del Norte Fundación Universitaria. 2026.

171 páginas

ISBN Físico: 978-958-8776-95-8

ISBN Digital: 978-958-8776-94-1

Palabras clave: Revisiones sistemáticas, Metanálisis, Epidemiología, Prácticas basadas en evidencia, Investigación en salud.

CDD 610.7

© Fundación Universitaria Católica del Norte

ISBN Físico: 978-958-8776-95-8

ISBN Digital: 978-958-8776-94-1

Fecha de edición: junio de 2026

Autores: Luis Felipe Higueta-Gutiérrez y Jaiberth Antonio Cardona-Arias

Edición: Editorial Católica del Norte Fundación Universitaria.

Corrección de estilo: Diego Alejandro Pérez Múnera

Diagramación y diseño: Alexander Sierra Gallego

Puede reproducirse con fines académicos citando la fuente respectiva.



Esta obra está bajo una licencia internacional Creative Commons Atribución-NoComercial-SinDerivadas 4.0.

Hecho en Colombia / Made in Colombia

Publicación financiada por Universidad de Antioquia.

Los autores son responsables, moral y legalmente, de la información presentada en este libro, así como del respeto a los derechos de autor; por tanto, no compromete en ningún sentido a la Fundación Universitaria Católica del Norte. Además, declaran la inexistencia de conflictos de interés de cualquier índole con instituciones o asociaciones comerciales.

Fundación Universitaria Católica del Norte // Cibercolegio UCN // Centro Laborem UCN Carrera 21 No. 34B-07; Santa Rosa de Osos (Antioquia, Colombia)

PBX: (60) (4) 605 15 35

Calle 52 No. 47-42; Medellín, Edificio Coltejer, piso 5 (Antioquia-Colombia)

PBX: (60) (4) 605 15 35

Portal web: www.ucn.edu.co / www.cibercolegioucn.edu.co

Tabla de contenido

Glosario	5
Introducción.....	10
En el principio...orígenes de la síntesis de la literatura científica	16
Capítulo 1. El diagnóstico: síntesis de la literatura de estudios que evalúan pruebas diagnósticas	26
Capítulo 2. La magnitud y distribución de una enfermedad: síntesis de la literatura de estudios descriptivos	60
Capítulo 3. Etiología de la enfermedad: síntesis de la literatura de estudios de casos y controles.....	89
Capítulo 4. Etiología de la enfermedad: síntesis de la literatura de estudios de cohorte	119
Capítulo 5. El tratamiento: síntesis de la literatura de estudios de intervención.....	143
Conclusiones y recomendaciones.....	166
Herramientas para realizar revisiones sistemáticas y metanálisis	169

Glosario

Tipos de revisiones

Narrativas: Síntesis de la literatura cuyo centro es un tema general y no una pregunta de investigación; por ello, carecen de metodología.

Panorámicas: Síntesis de la literatura orientada por preguntas de investigación amplias, cuyo centro son los atributos generales de los estudios incluidos. Están orientadas por una pregunta de investigación y tiene una metodología explícita predefinida y rigurosa para la búsqueda, selección y evaluación de los estudios a incluir; es decir, cumplen todos los requisitos de una revisión sistemática, solo se diferencian en el alcance amplio de la pregunta. Incluyen las revisiones de alcance (*scoping*, por sus siglas en inglés), de mapeo (*mapping*, por sus siglas en inglés) de brechas (*evidence gap maps*, por sus siglas en Inglés), entre otras.

Sistemáticas: Síntesis de la investigación o evidencia disponible que utiliza una metodología explícita predefinida para identificar, seleccionar críticamente y sintetizar los resultados de estudios primarios. Se caracterizan por proporcionar un análisis reproducible de la evidencia, responder a una pregunta de investigación específica, en la cual es fácilmente identificable el tipo de estudios y el desenlace.

De otras revisiones sistemáticas: son revisiones sistemáticas cuya unidad de análisis son revisiones sistemáticas; por eso, también se denominan meta-revisiones o revisiones sombrilla (*umbrella review*, por sus siglas en inglés).

Preguntas PICO

PICO: Población Intervención Comparador Desenlace (*Outcome*, por sus siglas en inglés), para estudios de intervención.

PICORT: Población Intervención Comparador Desenlace Recursos y Tiempo, para estudios de evaluación económica.

PEO: Población Exposición Desenlace, para estudios observacionales.

PICo: Población-Problema, Interés y Contexto, para revisiones de enfoque amplio o revisiones sistemáticas panorámicas.

Tipos de calidad

Editorial: Estándares de excelencia de los componentes de un artículo científico o de un texto en general (depende de la calidad metodológica o uno de sus componentes es la calidad metodológica).

Metodológica: Grado de validez interna (medir sin sesgos) y externa (generalización de los resultados) de un artículo.

Estadística: Claridad en la definición de las variables de un estudio y su medición, así como el uso de las pruebas estadísticas adecuadas para medir posibles variables de confusión, asociaciones, predicciones (según el tipo de estudio).

Metanálisis: Es un tipo de investigación que evalúa la calidad de la evidencia y sintetiza en una medida única los resultados de estudios originales que evalúan los mismos factores de exposición o desenlace. La búsqueda de los estudios originales debe ser sistemática, exhaustiva y reproducible, y, dependiendo del tipo de estudio que busquen sintetizar, se encuentran metanálisis de estudios observacionales, experimentales (clínicos o pre-clínicos), de pruebas diagnósticas.

Tipos de metanálisis

De efectos fijos: Método estadístico para combinar los resultados de estudios individuales que responde a la misma pregunta de investigación, asumiendo homogeneidad en el efecto a medir (no hay una elevada varianza entre los estudios) y que la variabilidad intra-estudios se explica por azar. En la estimación del intervalo de confianza de la medida global solo incluye la variabilidad intra-estudios.

De efectos aleatorios: Método estadístico para combinar los resultados de estudios individuales que responde a la misma pregunta de investigación, asumiendo heterogeneidad en el efecto a medir (hay una elevada varianza en el efecto entre los estudios). En la estimación del intervalo de confianza de la medida global incluye tanto la variabilidad intra-estudios como la variabilidad inter-estudios.

En red: Tipo de metanálisis que compara simultáneamente tres o más intervenciones en el cual no se incluyen simultáneamente en los ensayos clínicos incluidos en una revisión sistemática. Usa evidencia directa (de ensayos que comparan directamente las intervenciones de interés) e indirecta (derivada de ensayos

que comparan cada intervención con un comparador común o diferente). Este tipo de metanálisis en red se pueden clasificar en:

De medidas o comparaciones directas: Metanálisis en red con estudios que analizan las mismas intervenciones contra los mismos comparadores.

De medidas o comparaciones directas con diferentes comparadores: Metanálisis en red con estudios que contrastan la eficacia o efectividad de las intervenciones de interés, pero con diferencias en los comparadores de los estudios incluidos.

De comparaciones de tratamientos mixtos: Sinónimo de metanálisis en red que usa comparaciones de evidencia directa e indirecta.

Bayesianos: Combinación de evidencia de múltiples estudios mediante inferencia bayesiana. En la actualidad predominan dos enfoques para realizar inferencia estadística, el frecuentista y el Bayesiano. El primero se basa en los valores p y los intervalos de confianza, en los cuales se asume que los parámetros de un modelo son constantes y se pueden estimar a partir de los datos observados. Por su parte, el enfoque Bayesiano asume los parámetros como variables aleatorias, agrega probabilidades previas (de estudios previos o del concepto de expertos) sobre los parámetros de interés y las ajusta con datos observados de los estudios incluidos, lo que permite mejorar las estimaciones de probabilidades finales con cada nueva evidencia científica.

Prospectivos: Metanálisis en los cuales la selección de estudios se planifica antes de conocer los resultados de las investigaciones individuales, generalmente son colaboraciones entre centros de investigación que desarrollan estudios individuales y comparten datos para el análisis agrupado.

Tipos de desenlaces:

Duros: Eventos clínicos objetivos, bien definidos, que tienen un impacto directo en la salud de los pacientes, generalmente indicadores de morbilidad, mortalidad o discapacidad (en los tiempos recientes también se incluye la calidad de vida, siempre y cuando sea medida con constructos psicométricos robustos).

Blandos: Mediciones subjetivas o de salud percibida (por el componente perceptivo, algunos autores incluyen la calidad de vida).

Intermedios: Marcadores o mediciones del horizonte clínico que anteceden o predicen desenlaces duros y son clínicamente rele-

vantes.

Subrogados: Corresponde a un subtipo de desenlaces intermedios que se utilizan como sustitutos de los desenlaces clínicamente y tienen alta correlación con el desenlace duro.

Tipos de estudio epidemiológicos

Observacional: Estudios en los cuales los investigadores miden variables, sin controlar ninguna, para registrar datos del evento de estudio en su forma natural, y así establecer la magnitud y distribución de un evento (estudios descriptivos transversales) o factores de riesgo (analíticos o longitudinales).

Observacional descriptivo: Su objetivo principal es determinar la magnitud del evento y su distribución, usando la descripción básica de la epidemiología; es decir, con variables de tiempo, lugar y persona. Generalmente estudian solo un grupo de individuos.

Observacional analítico: Su objetivo es determinar relaciones o hacer comparaciones entre dos o más grupos, pueden ser transversales (conocido también como de corte o *cross sectional*, por sus siglas en inglés), longitudinales retrospectivos (el ejemplo clásico son los estudios de casos y controles) o longitudinales prospectivos (el ejemplo clásico son los estudios de cohorte).

De casos y controles: Su objetivo principal es identificar factores de riesgo; conforman al menos dos grupos, uno con evento y otros sin él, para evaluar de manera retrospectiva las variables independientes o potenciales factores de riesgo.

De cohortes: Su objetivo principal es identificar factores de riesgo mediante el seguimiento de al menos dos grupos (cohortes), uno expuesto y otro no expuesto, para evaluar de manera prospectiva las variables dependientes.

Estudio transversal: Mide la variable dependiente y las independientes en el mismo momento.

Estudio longitudinal: Mide la variable dependiente y las independientes en momento diferentes, se dividen en retrospectivos y prospectivos.

Experimental: Estudios en los cuales los investigadores manipulan una o más variables independientes (generalmente la asignación aleatoria a un grupo tratamiento o control) para evaluar el efecto (eficacia, efectividad o eficiencia) de una intervención, o

profundizar en el estudio de relaciones causa-efecto.

Ensayo clínico controlado: Su objetivo es evaluar la eficacia y seguridad de una intervención. Dado que se hace en condiciones controladas su desenlace central es la eficacia (la efectividad es el logro del resultado en condiciones no controladas o de la práctica clínica real-habitual). Se clasifican en fases I (datos preliminares del fármaco o vacuna como la farmacocinética, inmunogenicidad, dosis, etc.), II (seguridad y respuesta terapéutica inicial), III (eficacia y seguridad) y IV (farmacovigilancia).

Ensayo pragmático: Ensayo clínico con selección aleatoria que se desarrolla en las condiciones habituales de la práctica clínica o mundo real para evaluar efectividad de la intervención.

Ensayo comunitario: Similar al ensayo clínico, pero la unidad de aleatorización no es un individuo, sino una comunidad completa. Usados en los casos que no es imposible individualizar la intervención; por ejemplo, en temas de salud pública, salud ambiental y otras.

Cuantificación del efecto: Es la magnitud de la diferencia de una medida epidemiológica o estadística entre dos grupos; es decir, mide la fuerza de la asociación mediante diferencia de medias, diferencia de proporciones o razones, para mostrar la importancia clínica o epidemiológica del resultado (a diferencia de los valores p que solo indican la probabilidad de que un resultado no se explique por el azar).

Razón de prevalencia: prevalencia del evento en el grupo o subgrupo 1 dividida por la prevalencia del evento entre el grupo dos. Esta medida es usada en estudios transversales.

Razón de odds: Odds del evento en los expuestos dividida la odds del evento en los no expuestos. Esta medida es usada en estudios transversales en los que la frecuencia del evento es baja (usualmente menor al 20 %) y en estudios de casos y controles.

Riesgo relativo: Proporción de incidencia del evento en los expuestos, dividida la proporción de incidencia del evento en los no expuestos. Esta medida es usada en estudios longitudinales.

Razón de tasas o Hazard ratios: Compara la tasa de inciden-

cia de un evento entre dos grupos a lo largo del tiempo. La clave de esta medida es que tiene en cuenta, además del evento, el tiempo que tarda en generarse. Esta medida es usada en estudios longitudinales.

Introducción

*El grado de velocidad es directamente proporcional
a la intensidad del olvido (...) Nuestra época se entrega
al demonio de la velocidad y por eso se olvida
tan fácilmente a sí misma.
Milan Kundera. La lentitud.*

De acuerdo con el filósofo alemán Han (2022), vivimos en un *régimen de información*, una época en la que la información y su uso determinan sustancialmente los procesos políticos y económicos de las naciones, pero también moldea las decisiones domésticas y personales. El régimen de información opera apoderándose de la psique de las personas a través de un ensamblaje de dispositivos, que incluye las redes sociales, el internet, el teléfono celular, los algoritmos y, en general, los medios de comunicación digital. Entre más uso hagamos de estos dispositivos, más datos generamos.

Para hacer uso de esos grandes volúmenes de información se recurre al *big data* y la inteligencia artificial, herramientas que analizan la información, construyen perfiles con los datos analizados y explotan nuestras emociones, deseos e instintos con el propósito de influenciar nuestros comportamientos. Por esa razón, la información se ha convertido en uno de los bienes más valiosos de la economía contemporánea, lo que ha dado lugar a una suerte de frenesí informativo en el que circula rápidamente grandes volúmenes de información y, recurriendo a la terminología microbiológica, se viraliza, muta (tiene intervalo de actualidad reducido) y coloniza el pensamiento, desplazando procesos cognitivos como el saber, la racionalidad y el conocimiento.

En el ámbito científico, también asistimos a un tiempo en el que la producción de información se mueve a una velocidad difícil de aprehender, una velocidad excesivamente mayor a la de cualquier proceso académico de enseñanza-aprendizaje. Es difícil no inquietarse con estimaciones que muestran cómo en un día se publican

cerca de diez mil manuscritos o que, solo tomando PubMed, la cantidad de artículos en áreas biomédicas es de cinco mil textos por día (Desai et al., 2018).

La aceleración de la producción de información científica es un asunto que se alinea con los rasgos de esta época, descrita desde la sociología o la filosofía de Hartmut Rosa y Paul Virilio como la *sociedad de la velocidad o la aceleración* (Rosa, 2011), caracterizada, entre otros rasgos, por el rápido desarrollo de avances tecnológicos, de cambios socioculturales, de convulsiones políticas o de presiones económicas que exacerban los sentimientos de urgencia, de inmediatez, de eficiencia, entre otros asuntos que pueden generar una alta carga emocional y cognitiva para los sujetos; pero que simultáneamente los absorbe-atrapa por su eficiencia, por optimizar tareas cotidianas, por ofrecer mayor conectividad, por la facilidad para las interacciones sociales, entre otros beneficios.

En este balance de aspectos positivos y negativos de la *sociedad de la velocidad* (Rosa, 2011), es importante reflexionar sobre algunos de sus atributos diferenciadores: **i)** la percepción de que la cotidianidad está en un constante y acelerado cambio; **ii)** algunas dimensiones vitales que cada vez se descuidan más; **iii)** cada vez se reflexiona menos sobre la velocidad y las causas-consecuencias de su poder transformador en las personas, las instituciones, la democracia, la sociedad en general; y **iv)** en términos descriptivos y normativos se han identificado los efectos de este tipo de sociedad en el bienestar y la ética, pero ha sido un asunto de filósofos o sociólogos que no explicitan impactos en la esfera científica.

La aceleración de la vida científica y la alta producción de información se puede reflejar en algunas cifras de las principales bases de datos. En PubMed, como principal referente del campo biomédico, se publican entre 3.000 y 5.000 artículos por día, que sumado a otros campos científicos llega a cerca de 10.000, con tendencias de crecimiento exponencial (Desai et al., 2018). Es tal el incremento de la producción científica, que en la actualidad se dispone de un número igualmente creciente de revisiones sistemáticas; por ejemplo, en MEDLINE **se estima que se publicaron alrededor de 2.600 revisiones sistemáticas en el 2004, cifra que aumentó a 8.000 en el 2014, 21.000 en el 2018 y casi 30.000 en el 2019** (Hoffmann et al., 2021).

Esto puede implicar consecuencias positivas como la democratización de la ciencia, la toma de decisiones públicas con mejores evidencias, la orientación de prácticas disciplinarias con el mejor

conocimiento disponible, entre otros. Pero también subsume riesgos o efectos negativos como las crisis de la reproducibilidad de los estudios (Baker, 2016), la disminución de la calidad por la obsesión con la cantidad, las dificultades para separar la mejor evidencia en un mar de publicaciones, o que se produzca una cantidad inmanejable de estudios que nunca se usarán o que solo serán conocidos por sus autores.

En esta frontera, entre los impactos positivos y negativos de la velocidad y del régimen de la información en la producción científica, se halla una de las principales motivaciones de este libro, pensado para estudiantes jóvenes (principalmente) que se podrían beneficiar de esta aceleración científica. Además, partimos de una motivación clave para la elaboración de este texto, y fue la intención de optimizar procesos de enseñanza-aprendizaje en una época en la cual percibimos que se lee poco y, simultáneamente, se dispone de una cantidad inaprehensible de información y una cantidad de textos que no es posible conocer (es imposible pretender enseñar todo lo disponible en un área, incluso en la de mayor subespecialidad).

Las motivaciones para este libro también se encuentran en una concepción de la enseñanza que no gira en torno de la acumulación de contenidos, sino de la apropiación de métodos, estrategias o técnicas para agrupar elementos dispersos; asumiendo que la educación es un proceso continuo de síntesis de saberes heterogéneos, con énfasis en los métodos que permiten juzgar y apropiarse de lo más conveniente, lo cual aplica para el actuar científico donde no es tan determinante la acumulación de contenidos, sino la asimilación de los métodos.

También es claro que, entre los principales fines de la educación, está el escribir y leer bien, como reflejo de un pensamiento cada vez mejor ordenado y comunicado, para lo cual es determinante el buen uso de los métodos de las ciencias. En este caso, se ofrece un libro de texto que no solo orienta la realización de metanálisis en cada tipo de estudio clínico-epidemiológico, sino que también focaliza al lector en los aspectos clave de la divulgación de sus resultados.

Pasando al centro de los temas del libro, esta obra incluye una descripción de las características centrales de los principales diseños de investigación clínico-epidemiológicos y las etapas para desarrollar revisiones sistemáticas y metanálisis de este tipo de evidencia. Se conciben las revisiones sistemáticas como un tipo de investigación que busca, selecciona y sintetiza estudios publicados sobre una pregunta de investigación, cumpliendo los siguientes atributos: i)

sistematicidad, es decir, tener una pregunta y un protocolo *ex-ante* que oriente la ejecución del estudio; **ii**) exhaustividad o evidencia de que el protocolo de búsqueda está orientado a captar, reunir y analizar todos los estudios relacionados con la pregunta de investigación que orienta la revisión; **iii**) reproducibilidad de todas las fases de la investigación, pero con énfasis en la búsqueda y selección de los estudios, y en la extracción de las variables a analizar en ellos; y **iv**) evaluación de la calidad metodológica de los estudios incluidos¹ (Kronic Martinic et al., 2019)

Por su parte, los metanálisis corresponden a una serie de pruebas estadísticas que se deben utilizar en la estimación de las medidas combinadas que surgen de revisiones sistemáticas y que lograron incluir estudios similares en sus diseños y medición de variables de interés². Dichas pruebas estadísticas se dividen en el análisis de la varianza o el error estándar entre los estudios incluidos, la evaluación estadística del sesgo de publicación, el análisis de sensibilidad y la presentación de las mediciones individuales con la combinada por efectos fijos o aleatorios (Paul & Barari, 2022).

Por otra parte, también es importante resaltar que los textos de docencia disponibles en el campo de las revisiones sistemáticas han seguido la línea de Cochrane (Higgins et al., 2024), orientada principalmente al contexto de la evaluación de intervenciones y elaboración de guías de práctica clínica, o las recomendaciones de Instituto Joanna Briggs (Aromataris et al., 2024) focalizadas en sistematizar la evidencia para orientar decisiones clínicas o de política pública. Ambas instituciones están dirigidas a público con algún nivel de experticia, sin abordar necesidades de estudiantes o personas sin formación en epidemiología, que pueden usar estos métodos para diversos contextos académicos e investigativos.

En esta línea, se hace necesario un texto que ofrezca la experiencia de profesores e investigadores en este campo, para orientar a estudiantes e investigadores noveles en el desarrollo y lectura crítica de revisiones sistemáticas, metanálisis o simplemente en la síntesis

¹ Hasta hace algunos años, la calidad metodológica no era un atributo esencial para la definición de las revisiones sistemáticas (RS), sino que se asumía como una etapa del proceso de ejecución, pero no un elemento consustancial a la definición. Por ejemplo, para las RS con enfoque o alcance amplio de Cochrane, al ser estudios panorámicos sobre un tema que podría incluir varios tipos de estudio, varios desenlaces, o datos muy generales, no era obligatoria la evaluación de la calidad metodológica.

Pero en la actualidad, con el auge de las revisiones panorámicas (de mapeo, de alcance, exploratorias, entre otras), este cuarto numeral se está asumiendo como un elemento distintivo de la definición o diferenciación de las RS. De esta forma, las RS serían los estudios restringidos a un tipo de estudio que permita la evaluación de su calidad metodológica, mientras que las revisiones panorámicas no incluirían dicha evaluación de la calidad, cumplirían los demás atributos de la definición de revisiones sistemáticas (RS) pero no restringirían el tipo de estudio, de población o desenlace a medir.

² En décadas anteriores algunos editores de revistas biomédicas asumían el metanálisis como un tipo de estudio. En la actualidad, existe consenso en que el tipo de estudio es la revisión sistemática y el metanálisis un grupo de estadísticos para estimar heterogeneidad, sensibilidad, sesgos de publicación y medidas combinados.

de la literatura para responder a una pregunta de investigación. Así surge este libro, con el objetivo de describir las revisiones sistemáticas de la literatura y los metanálisis de los estudios epidemiológicos, enunciando sus etapas, los criterios de evaluación de la calidad y las alternativas analíticas que ofrece cada una de ellas. Estos contenidos son útiles para la orientación o mejoramiento de las preguntas de investigación de los estudiantes, la apropiación de los métodos de investigación o simplemente mejorar el estudio-acercamiento a sus disciplinas.

También, se espera que este libro sea útil para evitar que la sociedad de la velocidad colonice el diseño y ejecución de las revisiones sistemáticas y se llegue a una suerte de superficialidad, reduciendo este método a la aplicación de listas de chequeo (PRISMA para el protocolo de búsqueda y selección, EQUATOR para evaluar la calidad metodológica, etc.) o a la acción instrumental de aplicaciones-software (RAYYAN, RevMan, etc.), opacando el tiempo para razonar (la racionalidad requiere tiempo) sobre las definiciones que subyacen a los ítem o criterios de esta modalidad investigativa, sin un conocimiento claro sobre el tipo de problemas que resuelve, las exigencias de rigor o aspectos diferenciales para definir su validez externa, validez interna y exigencias de métodos estadísticos.

Finalmente, es importante citar algunos retos que motivaron la generación de este libro: la alta cantidad de publicaciones que demandan conocimientos sobre métodos de síntesis, los manuales disponibles para sintetizar evidencia que están dirigidos a población experta o con alta formación en el campo, se han reportado usos o aplicaciones cuestionables de las revisiones sistemáticas, y se dispone de pocos textos que aborden simultáneamente los conceptos y métodos epidemiológicos que subyacen a la síntesis con la generación de este tipo de evidencia meta-analítica. Desde la perspectiva docente y estudiantil, los textos están atomizados en algunas fuentes que profundizan en los diseños epidemiológicos, otros detallan aspectos estadísticos y en otros se presentan aspectos generales de metanálisis, sin disponer de un texto que una estos campos para optimizar o incluso simplificar los procesos de enseñanza-aprendizaje y le permita al estudiante tener una formación de alta calidad para evaluar críticamente la producción académica sobre los usos de los principales diseños epidemiológicos en sus temas de investigación.

Referencias

- Aromataris, E., Lockwood, C., Porritt, K., Pilla, B., & Jordan, Z. (Eds). (2024). *JBIManual for Evidence Synthesis* [Manual del JBI para la síntesis de evidencia]. JBI. <https://synthesismanual.jbi.global>. <https://doi.org/10.46658/JBIMES-24-01>
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility [1.500 científicos levantan la tapa sobre la reproducibilidad]. *Nature*, 533(7604), 452-454. <https://doi.org/10.1038/533452a>
- Desai, P., Telis, N., Lehmann, B., Bettinger, K., Pritchard, J. K., & Datta, S. (2018). *SciReader*: A Cloud-based Recommender System for Biomedical Literature* [SciReader*: Un sistema de recomendación basado en la nube para literatura biomédica]. bioRxiv. <https://web.stanford.edu/group/pritchardlab/publications/pdfs/Desai18-biorxiv.pdf>
- Han, B.-C. (2022). *Infocracia: la digitalización y la crisis de la democracia*. Taurus.
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds). (2024). *Cochrane Handbook for Systematic Reviews of Interventions version 6.5* [Manual Cochrane para Revisiones Sistemáticas de Intervenciones versión 6.5]. Cochrane.
- Hoffmann, F., Allers, K., Rombey, T., Helbach, J., Hoffmann, A., Mathes, T., & Pieper, D. (2021). Nearly 80 systematic reviews were published each day: Observational study on trends in epidemiology and reporting over the years 2000-2019 [Se publicaron casi 80 revisiones sistemáticas cada día: Estudio observacional sobre tendencias en epidemiología y presentación de informes durante los años 2000-2019]. *Journal of Clinical Epidemiology*, 138, 1-11. <https://doi.org/10.1016/j.jclinepi.2021.05.022>
- Krnic Martinic, M., Pieper, D., Glatt, A., & Puljak, L. (2019). Definition of a systematic review used in overviews of systematic reviews, meta-epidemiological studies and textbooks [Definición de revisión sistemática utilizada en revisiones generales de revisiones sistemáticas, estudios metaepidemiológicos y libros de texto.]. *BMC Medical Research Methodology*, 19(1), Article 203. <https://doi.org/10.1186/s12874-019-0855-0>

Paul, J., & Barari, M. (2022). Meta-analysis and traditional systematic literature reviews—What, why, when, where, and how? [Metanálisis y revisiones sistemáticas tradicionales de la literatura: ¿qué, por qué, cuándo, dónde y cómo?] *Psychol Mark*, 39(6), 1099-1115. <https://doi.org/10.1002/mar.21657>

Rosa, H. (2011). Aceleración social: consecuencias éticas y políticas de una sociedad de alta velocidad desentronizada. *Persona y Sociedad*, 25(1), 9-49. <https://doi.org/10.53689/pys.v25i1.204>

En el principio...orígenes de la síntesis de la literatura científica

Las revistas biomédicas tienen un impacto importante en la historia de la medicina, el ejercicio profesional, la docencia, la investigación y los progresos científicos. En la actualidad, es poco probable que existan países que no cuenten con por lo menos una revista biomédica, y, en ese sentido, una búsqueda rápida en la web sobre un tema de investigación específico arroja miles de resultados (Reyes, 2018).

No obstante, lo anterior, no siempre ha sido así. A comienzos del siglo XV, los conocimientos científicos eran divulgados de manera verbal en pequeñas reuniones de grupos de académicos, médicos o en las universidades. En los casos en los que tales reuniones no eran posibles por limitaciones geográficas, se intercambiaban cartas personales cuando se querían compartir los hallazgos o las opiniones sobre algunas observaciones clínicas, así como los avances sobre un tópico de investigación particular. Ese tipo de comunicación dificultaba el progreso científico por su lentitud, además permitía que algunos receptores de las comunicaciones cayeran en conductas no éticas y presentaran los hallazgos de sus colegas como propios (Reyes, 2018).

Este escenario cambió radicalmente con la invención de la imprenta a mediados del siglo XV, en cuanto fue posible la producción en serie de libros, y el alcance de lectores fue más amplio. Sin embargo, el libro tampoco era adecuado para divulgar los resultados de los nuevos conocimientos porque era necesario reunir un alto

volumen de contenido para justificar su elaboración. Así que varios investigadores optaron por divulgar sus conocimientos a través de folletos o panfletos que ulteriormente evolucionarían al artículo científico (Jiménez Villa et al., 2010).

La historia del artículo científico está estrechamente relacionada con las actividades de la Royal Society de Londres. En 1665, esta sociedad inicia la publicación de *The Philosophical Transactions*, el primer periódico científico, cuyo contenido giraba entorno de noticias, cartas y algunos informes experimentales de interés para la ciencia (Marta, 2015). A inicios del siglo XVIII, *The Philosophical Transactions* inició con las publicaciones sobre temas médicos, predominantemente en latín y sin muchas contribuciones clínicas importantes. En un intento por mejorar la calidad de sus publicaciones, a mediados del siglo XVIII se creó un comité para la revisión de todos los documentos previo a su publicación, proceso que hoy se conoce como la revisión por pares (Marta, 2015).

Además de *The Philosophical Transactions*, la historia de las revistas médicas está relacionada con la Escuela de Medicina de Edimburgo y sus publicaciones médicas iniciadas a partir de 1731, publicaciones que luego se convirtieron en el *Edinburgh Medical Journal* (Marta, 2015).

A finales del siglo XVIII, las revistas ya habían reemplazado a los libros como medio de divulgación de los hallazgos de observaciones clínicas y de las investigaciones; sin embargo, solo hasta inicio del siglo XIX se empezaron a publicar de forma permanente revistas médicas de alta calidad. Así, *The New England Journal of Medicine (NEJM)* se inicia en 1812; *The Lancet* apareció por primera vez en 1823, con el propósito de publicar las conferencias de médicos que trabajaban en las escuelas de medicina de Londres y de promover informes de casos escritos por las autoridades médicas y quirúrgicas de la época; más tarde, en 1857, se publica por primera vez el *British Medical Journal (BMJ)* (Jiménez Villa et al., 2010; Marta, 2015).

El periodo restante del siglo XIX y principios del siglo XX fue testigo de la aparición de revistas especializadas en diferentes campos como fisiología, patología, bacteriología, medicina tropical, cirugía, entre otros (Marta, 2015; Reyes, 2018). A partir de entonces, se han fundado miles de revistas médicas y con ellas han aparecido sistemas de clasificación en bases de datos bibliográficas que permiten diferenciar a las “revistas de corriente principal” (las que tienen mayor número de citas) y las de la “periferia científica”. Las revistas científicas más prestigiosas están incluidas en bases de da-

tos mundialmente reconocidas como: MEDLINE/ PubMed, Science Citation Index, Scopus y/o SciELO (Reyes, 2018).

En este sentido, también es posible encontrar rankings para las revistas, entre ellos se destaca el *Scimago Journal & Country Rank* (<https://www.scimagojr.com>), una página en internet que genera indicadores sobre el impacto y la calidad de las publicaciones. En este ranking, se clasifican las revistas por cuartiles Q1 (Verde), Q2 (Amarillo), Q3 (Naranja) y Q4 (Rojo), áreas de conocimiento y país; así, las revistas que se ubican en el primer cuartil son aquellas que se cuentan entre el 25 % de las revistas más prestigiosas del mundo.

La historia de las revistas científicas en Colombia, hasta la primera mitad de 1990, se resume en esfuerzos de iniciativas de las universidades y algunas organizaciones independientes por divulgar las investigaciones de sus profesores e investigadores y dar cuenta de su vida académica e intelectual. El objetivo era condensar, en un único número al año, la producción de sus investigadores (Suárez Guava, 2017). Finalizando 1990 la situación empieza a cambiar con la creación de Publindex y sus requisitos de indización que incluían la periodicidad (número de veces que se publica al año) y demostrar que los artículos sometidos pasaban por un proceso de evaluación por pares anónimos. Así, se inició un sistema de clasificación de las revistas nacionales en A1, A2, B o C, dependiendo de la calidad y de los manuscritos publicados (Suárez Guava, 2017).

Con el tiempo, el sistema de clasificación de las revistas también fue asumido como indicador indirecto de la calidad de las instituciones a las que pertenecían, de manera que un gran número de universidades dedicó esfuerzos al fortalecimiento de sus revistas científicas o a la creación de nuevos medios de divulgación. Una muestra de lo anterior se encuentra en que para el año 1996 había 29 revistas científicas en Publindex; en el año 2004 había 119 revistas indizadas y para el año 2016 la cifra alcanzaba las 526 publicaciones (Min-ciencias, 2025). Así las cosas, en menos de 20 años la cantidad de revistas reconocidas por el sistema de indexación nacional creció en más de 1.800 por ciento (Suárez Guava, 2017).

Finalizando el año 2016, Publindex realiza otra modificación importante en su modelo de indización, al incorporar el nivel de citación como criterio para evaluar el impacto de las publicaciones (Departamento Administrativo de Ciencia Tecnología e Innovación, 2016). De acuerdo con el nuevo modelo, el 53 % de las revistas indizadas desapareció y en el 2017 el número de revistas científicas en el país pasó a ser de 246. Los datos del 2020 ponen de manifiesto

que el país cuenta con 275 revistas indizadas en Publiindex, y el 15 % de ellas (n=40) en tópicos relacionados con ciencias médicas y de la salud (Minciencias, 2025). De esas revistas, 12 están clasificadas en la categoría B y 28 en la categoría C. Con respecto a la clasificación de las revistas nacionales en el ranking internacional de Scimago, específicamente en la categoría de medicina, se encuentran 8 publicaciones, dos de ellas en la categoría Q3 (Colombia médica y biomédica) y las demás en la categoría Q4.

Con base en lo expuesto en los párrafos previos, se puede inferir que uno de los desafíos a los que se enfrenta quien desee emprender una investigación en el ámbito de la salud es encontrar la información de interés y seleccionar la más relevante en medio de la enorme cantidad disponible en la literatura científica. En este sentido, **en 1994 Mulrow estimó que cada año se publicaban más de dos millones de artículos en áreas biomédicas y dibujó una imagen mental de una torre de 500 metros de altura si todas las revistas se ubicaran unas sobre otras** (Mulrow, 1994).

En la actualidad, el crecimiento en el número de artículos se ha acelerado y se estima que las cifras alcanzan los dos millones al año (Clarke, 2018). Esto converge con otros cálculos que han evidenciado la exagerada velocidad con la cual ha crecido la producción del conocimiento académico en la historia de la humanidad; tomando como una unidad la cantidad de textos acumulados hasta el año uno de la era cristiana, la humanidad demoró casi 16 siglos para duplicar esta cifra; la segunda duplicación tomó 250 años (año 1500-1750), la tercera 150 (año 1900), y en el presente se estima que la duplicación del conocimiento acumulado solo demora 1-2 años (Ruiz-Healy, 2011).

Lo enunciado se podría exponer de manera paralela con cambios paradigmáticos que, cada vez, ocurren en periodos más cortos de tiempo. Por ejemplo, Ray Kurzweil (1999), en el libro *“La Edad de las máquinas espirituales”*, indica que hasta el año mil de la era cristiana los cambios de paradigmas demoraban miles de años, durante el segundo milenio tardaban cerca de un siglo, en el siglo XIX se presentaron más cambios paradigmáticos que en toda la historia previa, en el siglo XX demoraban cerca de una década, y en la actualidad el tiempo de cambio en los paradigmas predominantes es aún menor.

Por supuesto, el incremento de las publicaciones científicas es deseable por su contribución al progreso de la ciencia, la divulgación de los nuevos conocimientos a un público cada vez más amplio y la garantía que ofrecen sobre la propiedad intelectual y la protección

frente al plagio (Reyes, 2018). En el campo de la medicina, las publicaciones biomédicas proporcionan los fundamentos para orientar la práctica clínica y mejorar la calidad de la atención a los pacientes, el sistema sanitario en general y focalizar la distribución de los recursos. También se publican muchos artículos para que los autores puedan ascender en la jerarquía médica o académica, lo que a su vez trae beneficios a las instituciones en las que trabajan, como mayor visibilidad o mejor financiación (Marta, 2015).

Pese a ello, con cierta regularidad se descubren casos de mala conducta científica. Según el número de casos confirmados por el Gobierno, en los Estados Unidos el fraude se documenta en aproximadamente 1 de cada 100.000 publicaciones o en 1 de cada 10.000, según la fuente (Fanelli, 2009). También se estima que entre 0,02 y 0,2 % de los trabajos en la literatura científica son fraudulentos. En el mismo sentido, algunas auditorías realizadas por la *Food And Drugs Administration* (FDA), entre 1977 y 1990, pusieron de manifiesto deficiencias entre el 10 % y el 20 % de los estudios, y llevaron al 2 % de los investigadores clínicos a ser declarados culpables de mala conducta científica (Fanelli, 2009).

Sumado a la mala conducta científica, una preocupación actual en la comunidad académica está relacionada con la reproducibilidad de las investigaciones; es decir, que cuando un estudio se repite en condiciones similares, no se obtienen los mismos resultados. La falta de reproducibilidad puede ocurrir por razones legítimas como la variabilidad natural en los sistemas biológicos; sin embargo, también pueden ser generadas por problemas metodológicos de las investigaciones debido a sesgos en los investigadores, los instrumentos de medición o los participantes, la baja potencia estadística de los análisis, la flexibilidad en el diseño de los estudios, entre otros factores (The Academy of Medical Sciences, 2015).

En adición a lo anterior, las décadas recientes se han caracterizado por la multiplicidad de incentivos económicos y de reconocimiento que han derivado en un sistema investigativo que privilegia la cantidad superficial sobre la calidad de los manuscritos, afectando procesos académicos-investigativos y la integridad científica.

El investigador que quiera orientar su práctica clínica con base en la evidencia se enfrenta a dos problemas importantes: por un lado, la abrumadora cantidad de información disponible en la web, producto de la proliferación de revistas científicas, y, por el otro, la baja calidad metodológica de algunas de las investigaciones publicadas. En este sentido, algunos autores han estimado que la mayoría

de artículos publicados solo son leídos por los autores, revisores y editores; que menos del 10 % son citados (tal vez sin leerse); muchas instituciones privilegian la publicación de temas mediáticos que no aportan a problemáticas reales o solo buscan impactos bibliométricos, llegando al punto de un elevado número de publicaciones con textos de calidad editorial y científica cuestionable (El Observatorio de la Universidad Colombiana, 2023).

En este contexto, las revisiones sistemáticas de la literatura proporcionan un recurso valioso para resolver ambos problemas, porque su metodología permite sintetizar la evidencia disponible sobre un tópico de investigación y evaluar la calidad de esas investigaciones. Dados esos atributos, una búsqueda en PubMed en abril de 2025 para artículos etiquetados con el término metanálisis recuperó más de 300 mil registros, con un crecimiento exponencial desde el 2010; al punto que solo en el 2024 esta base de datos publicó más de 23.000 artículos clasificados como metanálisis. Algunos autores han estimado que el total disponible en la literatura ya supera los 200.000 (Clarke, 2018).

Aunque las revisiones sistemáticas se consideran procesos científicos imparciales y desapasionados, se han documentado diferentes estrategias a través de las cuales se pueden introducir sesgos en ellas o en los estudios primarios objeto de la síntesis. Los investigadores toman decisiones subjetivas influyentes en todas las etapas de una revisión: formulación de la pregunta, la estrategia de búsqueda, la aplicación de criterios de inclusión y exclusión, la extracción de la información, los análisis estadísticos, además de los conflictos de interés. En vista de la importancia del rigor y la calidad metodológica que se requieren en las revisiones sistemáticas, se han publicado diferentes iniciativas para garantizar el rigor metodológico de esta modalidad de investigación; pese a ello, las críticas a la calidad de las revisiones sistemáticas son generalizadas (Uttley et al., 2023).

Hay muchos informes que indican que se está publicando una gran cantidad de revisiones sistemáticas y metanálisis innecesarios, engañosos, redundantes y de mala calidad. Algunos autores sostienen que, en lugar de promover la medicina y la atención de la salud basadas en la evidencia, esta modalidad de investigación se ha convertido en un proyecto de fácil producción y publicación o en herramientas de marketing manipulables por las industrias farmacéuticas, alimentarias, entre otras (Ioannidis, 2016).

Paralelo a la epidemia de revisiones sistemáticas de mala calidad, ha emergido una amplia variedad de tipologías; por

ejemplo, Sutton et al. (2019) identificaron 48 tipos de revisiones que se clasifican en siete familias definidas como: revisiones tradicionales, revisiones sistemáticas, revisiones de revisiones, revisiones rápidas, revisiones cualitativas, revisiones de métodos mixtos y revisiones de propósitos específicos. A su vez, descomponen las familias en revisiones integrativas, revisiones de estado del arte, revisiones sistemáticas de estudios epidemiológicos, umbrela reviews, metanálisis en red, revisiones realistas rápidas, meta-síntesis, entre otras (Sutton et al., 2019).

Lo expuesto pone al descubierto que en este campo de la investigación se ha generado una suerte de torre de babel en la que los investigadores, a su criterio, asignan nombres y apellidos a las revisiones sistemáticas, y agregan o eliminan etapas de manera caprichosa que desvirtúan el objetivo inicial de la síntesis de la literatura, lo que a su vez dificulta identificar los criterios de calidad de este tipo de estudios y, por tanto, juzgarlos.

Esta problemática ocurre a pesar de la disponibilidad de guías como las del instituto Cochrane o la del instituto Joanna Briggs. La primera es una institución que fue creada con el objetivo de contribuir a *un mundo de mejor salud donde las decisiones sobre salud y atención médica se basan en evidencia de investigación sintetizada, actualizada, relevante y de alta calidad* (West, 2016). Sin embargo, no ha estado exenta de polémica por publicación de revisiones sistemáticas sesgadas; en este sentido, el caso del estudio de la vacuna de VPH, que le costó la renuncia a Jørgensen y otros miembros de la colaboración Cochrane, resulta emblemático (Jørgensen et al., 2018; Sturmberg, 2019).

Sumado a ello, se critica la estructura jerárquica y cerrada de la organización que hace difícil contradecir las opiniones preconcebidas de los expertos internos y, por tanto, publican guías altamente especializadas que, en lugar de contribuir a democratizar el conocimiento, lo complejiza para que continúe centralizado en una élite académica (Sturmberg, 2019). Por otra parte, el instituto Joanna Briggs surge con el objetivo de mejorar los resultados de salud en comunidades de todo el mundo, mediante la promoción y el apoyo del uso de la mejor evidencia disponible para fundamentar las decisiones que se toman en los centros de atención en salud (JBI, 2025); sin embargo, las guías que ofrecen al público utilizan un lenguaje especializado dirigido a expertos en el campo o con conocimientos amplios en epidemiología, dejando por fuera buena parte de la población a la que dicen digerirse.

Este libro ofrece la experiencia de los dos autores en esta modalidad de investigación, con el objetivo de orientar a estudiantes e investigadores noveles del campo de la salud, en el desarrollo de revisiones sistemáticas, metanálisis o simplemente en la síntesis de la literatura para responder a una pregunta de investigación.

El libro está estructurado sobre la base del horizonte clínico de la enfermedad, partiendo del diagnóstico y finalizando con los tratamientos, pero también integra la lógica de la jerarquía de la evidencia que ubica los estudios sobre magnitud y distribución de la enfermedad en los primeros peldaños; va ascendiendo por los estudios sobre factores de riesgo y cierra con los estudios de intervención. Así las cosas, el lector encontrará los elementos que le permitan formular preguntas sobre síntesis de la literatura para la magnitud y distribución de una enfermedad, la etiología, los factores de riesgo y pronóstico, la utilidad de los procedimientos diagnósticos y la eficacia de distintas intervenciones terapéuticas o preventivas.

En los capítulos del uno al cinco se desagregan los asuntos conceptuales y metodológicos para la síntesis de la literatura en los siguientes diseños: pruebas diagnósticas, prevalencia y transversales, factores de riesgo o pronóstico y ensayos clínicos. Cada capítulo desarrolla las características distintivas del tipo de estudio, su validez interna, externa y sus potenciales sesgos; plantea escenarios sobre cuándo se debe sistematizar la evidencia del tipo de estudio abordado, genera recomendaciones para la elaboración del protocolo de búsqueda con sus criterios de inclusión y exclusión, la extracción de la información y su análisis; acompañando cada etapa con la interpretación y recomendaciones de los autores. El cierre de cada capítulo contiene un ejemplo práctico de un metanálisis en el tipo de estudio descrito, enunciando los tópicos centrales de la problematización, métodos, presentación de resultados y conclusión.

Referencias

Clarke, M. (2018). Partially systematic thoughts on the history of systematic reviews [Reflexiones parcialmente sistemáticas sobre la historia de las revisiones sistemáticas]. *Systematic reviews*, 7(1), Article 176. <https://doi.org/10.1186/s13643-018-0833-3>

Departamento Administrativo de Ciencia Tecnología e Innovación. (2016). *Política para mejorar la calidad de las publicaciones científicas nacionales* (Documento 1601). <https://minciencias.gov.co/sites/default/files/upload/paginas/politica-publindex-colciencias.pdf>

El Observatorio de la Universidad Colombiana. (2023). *¿Qué tanto debe ser leído y citado un artículo científico?: Un debate con impacto económico*. <https://www.universidad.edu.co/que-tanto-debe-ser-leido-y-citado-un-articulo-cientifico-un-debate-con-impacto-economico/>

Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data [¿Cuántos científicos inventan y falsifican investigaciones? Una revisión sistemática y un metanálisis de datos de encuestas.]. *PLoS ONE* 4(5), Article e5738. <https://doi.org/10.1371/journal.pone.0005738>

Ioannidis, J. P. (2016). The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses [La producción masiva de revisiones sistemáticas y metanálisis redundantes, engañosos y conflictivos]. *The Milbank quarterly*, 94(3), 485-514. <https://doi.org/10.1111/1468-0009.12210>

JBI. (2025). *Who are we?* [Quiénes somos nosotros] <https://jbi.global/about-jbi>

Jiménez Villa, J., Argimon Pallàs, J. M., Martín Zurro, A., & Vilardell Tarrés, V. (2010). *Publicación científica biomédica. Cómo escribir y publicar un artículo de investigación*. Elsevier.

Jørgensen, L., Gøtzsche, P. C., & Jefferson, T. (2018). The Cochrane HPV vaccine review was incomplete and ignored important evidence of bias [La revisión Cochrane de la vacuna contra el VPH fue incompleta e ignoró evidencia importante de sesgo]. *BMJ Evidence-Based Medicine*, 23(5), 165-168. <https://doi.org/10.1136/bmjebm-2018-111012>

Kurzweil R. (1999). *La Edad de las máquinas espirituales*. Planeta.

Marta, M. M. (2015). A brief history of the evolution of the medical research article. *Medicine and Pharmacy Reports*, 88(4), 567-570. <https://doi.org/10.15386/cjmed-560>

Minciencias. (2025). *La ciencia en cifras*. <https://minciencias.gov.co/la-ciencia-en-cifras/publindex/estadisticas-2004-2016>

- Mulrow, C.D. (1994). Rationale for systematic reviews [Justificación de las revisiones sistemáticas]. *BMJ*, 309(6954), 597-599. <https://doi.org/10.1136/bmj.309.6954.597>
- Reyes, B. H. (2018). Historia, propósitos y características de las revistas médicas. *Revista Médica de Chile*, 146(8), 913-920. <http://dx.doi.org/10.4067/s0034-98872018000800913>
- Ruiz-Healy E. (2011). *La aceleración del conocimiento*. <https://www.merca20.com/la-aceleracion-del-conocimiento/>
- Sturmberg J. P. (2019). Evidence-based medicine-Not a panacea for the problems of a complex adaptive world [La medicina basada en la evidencia no es una panacea para los problemas de un mundo complejo y adaptativo]. *Journal of Evaluation in Clinical Practice*, 25(5), 706-716. <https://doi.org/10.1111/jep.13122>
- Suárez Guava, L.A. (2017). Una breve historia de las revistas científicas en Colombia o la maldición de ser editor. *Universitas Humanística*, (83), 9-15. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-48072017000100009
- Sutton, A., Clowes, M., Preston, L., & Booth, A. (2019). Meeting the review family: exploring review types and associated information retrieval requirements [Conociendo la familia de revisiones: exploración de los tipos de revisiones y los requisitos de recuperación de información asociados]. *Health Information and Libraries Journal*, 36(3), 202-222. <https://doi.org/10.1111/hir.12276>
- The Academy of Medical Sciences. (2015). *Reproducibility and reliability of biomedical research* [Reproducibilidad y fiabilidad de la investigación biomédica]. <https://acmedsci.ac.uk/file-download/38189-56531416e2949.pdf>
- Uttley, L., Quintana, D. S., Montgomery, P., Carroll, C., Page, M. J., Falzon, L., Sutton, A., & Moher, D. (2023). The problems with systematic reviews: a living systematic review [Los problemas de las revisiones sistemáticas: una revisión sistemática viva]. *Journal of Clinical Epidemiology*, 156, 30-41. <https://doi.org/10.1016/j.jclinepi.2023.01.011>
- West, B. J. (2016). *Simplifying Complexity. Life is Uncertain, Unfair and Unequal* [Simplificando la complejidad. La vida es incierta, injusta y desigual]. Bentham e Books.

Capítulo 1. El diagnóstico: síntesis de la literatura de estudios que evalúan pruebas diagnósticas

1. Características de los estudios de pruebas diagnósticas

La evaluación de pruebas diagnósticas es un tipo de estudio que valora qué tan buena es una prueba para identificar poblaciones con presencia o ausencia de una enfermedad. Las pruebas diagnósticas pueden incluir el examen físico, psicológico, imagenológico (rayos X, tomografías, resonancias), pruebas de laboratorio (microbiológica, bioquímicas, inmunológicas) y su objetivo es distinguir entre salud y enfermedad, entre diferentes enfermedades con signos y síntomas similares o entre diferentes etapas de una misma enfermedad (Bouter et al., 2018).

Los resultados de una prueba diagnóstica tienen implicaciones importantes porque derivan en la realización de pruebas más sofisticadas, someterse a procedimientos invasivos, iniciar una terapia farmacológica, mayores costos, generan ansiedad y preocupación en los afectados e incluso pueden decidir sobre el acceso o no a un empleo (Celentano et al., 2024). Por tanto, la calidad del proceso diagnóstico es un tema fundamental para la clínica (médico-paciente), para la salud pública (prevención primaria y secundaria) y para la investigación epidemiológica (Celentano et al., 2024).

Las pruebas diagnósticas con mejor desempeño suelen ser costosas, invasivas, requieren de un tiempo considerable para ser procesadas, mayor y mejor infraestructura o de personal altamente calificado, de manera que no son adecuadas para establecerse como pruebas de rutina. En este sentido, en la práctica clínica se buscan pruebas diagnósticas que además de tener buen desempeño, sean rápidas, simples, económicas y no invasivas, así que una buena parte de la investigación diagnóstica se centra en buscar nuevas técnicas que cumplan con esas características (Bouter et al., 2018).

En coherencia con lo anterior, en la evaluación de una prueba diagnóstica se compara una prueba de interés (prueba índice) frente

a una prueba que se sabe que es la mejor disponible para identificar la presencia o ausencia de una enfermedad (prueba de referencia). Para ello, se desarrolla un estudio transversal analítico en el que se involucran todos los pacientes sospechosos de tener la enfermedad, y a los cuales se les aplica simultáneamente o en una ventana de tiempo que no afecte el resultado (en estados de salud que no cambien rápidamente en el tiempo) la prueba índice y la prueba de referencia; por último, se comparan los resultados de las dos pruebas para develar la capacidad que tiene la prueba índice para discriminar o diferenciar sanos de enfermos.

Para la evaluación diagnóstica se tienen en cuenta las características de validez, reproducibilidad y seguridad o desempeño (Campbell et al., 2017; Leeflang, 2014). **La validez de una prueba es la capacidad que esta tiene para diferenciar las personas sanas de las enfermas; se evalúa con los parámetros de sensibilidad y especificidad, y los índices que se derivan de estos dos parámetros**, como los cocientes-razones de probabilidad o verosimilitud, la curva ROC y la razón de odds diagnóstica (Pita Fernández & Pértegas Díaz, 2003).

En este orden de ideas, **la reproducibilidad o precisión hace referencia a la capacidad de la prueba para ofrecer los mismos resultados cuando se aplica en condiciones similares. La reproducibilidad se evalúa con la concordancia y los coeficientes de correlación**³. La seguridad indica la capacidad de la prueba de predecir la presencia o ausencia de enfermedad; además, se evalúa con los valores predictivos positivos y negativos (Pita Fernández & Pértegas Díaz, 2003), y en algunos contextos se denominan desempeño diagnóstico, dado que indican la probabilidad que una persona tenga o no una enfermedad por el resultado de la prueba.

La sensibilidad de una prueba diagnóstica (prueba índice) es definida como la probabilidad de que una persona con la enfermedad tenga un resultado positivo (Leeflang, 2014). Esta medida indica el porcentaje de los pacientes con la enfermedad de interés que son clasificados correctamente como enfermos por la prueba índice. La sensibilidad se expresa entonces como una proporción, en la cual el número de individuos con la enfermedad que dan positivo con la prueba está en el numerador (a) y el número total de individuos con la enfermedad está en el denominador (a + c) (Aschengrau & Sea-

³ Estos parámetros no son frecuentes en algunos estudios de evaluación de pruebas diagnósticas, dado que se acostumbra más en contextos de comparación de pruebas similares. En sentido estricto, en la evaluación de pruebas diagnósticas se parte del hecho que el estándar es mejor que la prueba índice y por ello son más importantes los índices de validez y seguridad.

ge, 2020). Así, si la prueba detecta a todos los pacientes que tienen la enfermedad (todos dan positivo) daría una sensibilidad máxima, es decir, del 100 % (en algunos casos también se expresa como probabilidad con valores entre 0 y 1). Una prueba altamente sensible, excluye una baja proporción de personas con la condición, lo que se interpreta como una baja proporción de falsos negativos ($c/a+c$).

En este punto es importante aclarar que el término sensibilidad (diagnóstica) se refiere a la situación clínica en la que se está utilizando una prueba, y es diferente de la sensibilidad analítica, la cual hace referencia a la mínima unidad de cambio de la medición que la prueba puede detectar (Leeflang, 2014). Además, el término “positivo” puede ser confuso, al asumirse que es un resultado favorable y concluir incorrectamente que no tiene la enfermedad (Bouter et al., 2018).

Por su parte, la especificidad es definida como la probabilidad de que una persona sin la enfermedad tenga un resultado negativo (Leeflang, 2014). La especificidad de una prueba indica el porcentaje de un grupo de personas sin la enfermedad que se clasifica correctamente como libre de ella por la prueba índice. De manera que la especificidad también se expresa como una proporción, en el que el número de individuos sin la enfermedad que dan negativo con la prueba índice está en el numerador (d), y el número total de individuos sin la enfermedad está en el denominador ($b + d$) (Aschengrau & Seage, 2020). Una especificidad óptima del 100 % se alcanza cuando la prueba índice da negativo en todos los pacientes libres de la enfermedad (Bouter et al., 2018). Una prueba altamente específica implica una baja proporción de personas sin la condición que se clasifiquen equívocamente como positivos, lo que se interpreta como una baja proporción de falsos positivos ($b/b+d$).

Por otro lado, también es necesario precisar que la especificidad diagnóstica es diferente de la especificidad analítica, ya que esta última se refiere a la especificidad del analítico o microorganismo a detectar (si no es específico se puede presentar reactividad cruzada); por ejemplo, la prueba para la detección de anticuerpos del VIH es específica porque identifica anticuerpos contra antígenos recombinados derivados de las regiones *Env* y *Pol* del VIH-1 y del VIH-2 (Leeflang, 2014).

La especificidad y la sensibilidad son propiedades intrínsecas de la técnica diagnóstica, por lo que no están afectadas por la prevalencia del evento en una población determinada, y por ello dan cuenta de la validez; es decir, estos parámetros confirman si la prueba

índice mide lo que pretende medir. Estos parámetros de validez se ven afectados por la etapa del horizonte clínico en que se apliquen, por ello es común que la evaluación de la sensibilidad y la especificidad se realicen en personas asintomáticas, sintomáticas y en estados avanzados del evento.

El valor predictivo positivo indica la probabilidad de que una persona con una prueba diagnóstica positiva tenga efectivamente la enfermedad (Bouter et al., 2018). En el entorno clínico, esta información es particularmente útil porque al médico le interesa saber, al recibir los resultados de una prueba diagnóstica positiva, cuál es la probabilidad de que el paciente que tiene enfrente tenga la enfermedad. Para calcular la proporción de pacientes que dan positivo en la prueba índice y que en realidad presentan la enfermedad en cuestión, se divide el número de verdaderos positivos (a) entre el número de pacientes que dieron positivo con la prueba índice (a + b) (Celentano et al., 2024).

Un punto que se debe tener en consideración con respecto a los valores predictivos positivos es que estos dependen en gran medida de la prevalencia de la enfermedad de interés, así cuanto mayor sea la prevalencia, mayor será el valor predictivo positivo y una prevalencia más baja dará como resultado un valor predictivo positivo más bajo. La relación entre la prevalencia de la enfermedad y el valor predictivo pone de manifiesto que los resultados de una técnica diagnóstica deben interpretarse teniendo en cuenta la frecuencia de la enfermedad en la población de la que proviene el paciente, ya que es posible que el mismo resultado de una prueba tenga un valor predictivo positivo muy diferente cuando se administra a un paciente que proviene de una población de alto riesgo (alta prevalencia) que cuando se aplica a un paciente que proviene de una población de bajo riesgo (baja prevalencia) (Celentano et al., 2024). En consecuencia, el resultado de la prueba por sí solo puede no ser suficiente para guiar el diagnóstico de la enfermedad (Celentano et al., 2024).

El valor predictivo negativo indica la probabilidad de que una persona con este resultado no tenga la enfermedad. En este caso, el médico se preguntará, al recibir un paciente con un resultado de un test negativo, cuál es la probabilidad de que ese paciente no presente la enfermedad. Este parámetro se calcula dividiendo los verdaderos negativos (d), entre todos aquellos que dieron negativo con la prueba índice (c + d). El valor predictivo negativo también se ve afectado por la prevalencia, pero en este caso una prevalencia más baja dará como resultado un valor predictivo negativo más alto, y

una prevalencia alta dará como resultado un valor predictivo negativo bajo (Bouter et al., 2018; Celentano et al., 2024).

La especificidad, la sensibilidad y los valores predictivos son los parámetros más utilizados al evaluar las propiedades de una técnica diagnóstica. Sin embargo, estos parámetros no permiten establecer el nivel óptimo de relación entre sensibilidad y especificidad; es decir, dejan al investigador en la necesidad de tomar una prueba por su mayor sensibilidad (una práctica pasada para los programas de tamización) o especificidad (para la confirmación diagnóstica). De la misma manera, en desenlaces continuos dejan al usuario en dilemas donde las ganancias en sensibilidad se correlacionan con pérdidas de especificidad, y viceversa. Por tanto, para dimensionar el beneficio clínico de una prueba resultan de mayor utilidad las razones de verosimilitud (*likelihood ratio*).

Las razones de verosimilitud se definen como la razón entre la posibilidad de obtener un resultado en los pacientes con la enfermedad y la posibilidad de obtener ese mismo hallazgo en individuos sin enfermedad (Fuente-Alba & Molina Villagra, 2017). Así, una razón de verosimilitud positiva indicaría la razón entre la proporción de positivos entre quienes tienen la enfermedad (sensibilidad) y los falsos positivos (1-especificidad). La razón de verosimilitud negativa refleja la razón entre la proporción de falsos negativos (1-sensibilidad) y la proporción de verdaderos negativos (especificidad) (Lee-flang, 2014).

Para ilustrar lo anterior, una razón de verosimilitud positiva de 10 indica que es 10 veces más verosímil que un individuo con la enfermedad obtenga una prueba positiva, que un individuo sin la enfermedad obtenga ese mismo resultado (o que de cada 10 verdaderos positivos que detecta la prueba índice se presenta 1 falso positivo con la misma). Así, una prueba diagnóstica será de mayor utilidad en la medida en que la razón de verosimilitud positiva sea mayor, puesto que permite confirmar la presencia de la enfermedad, y en la medida en la que la razón de verosimilitud negativa tenga un valor bajo, debido a que descarta la enfermedad (Fuente-Alba & Molina Villagra, 2017) (ver Tabla 1). Las razones de verosimilitud son particularmente útiles en la toma de decisiones clínicas porque son valores inherentes a la prueba diagnóstica e independientes de la prevalencia de la enfermedad, y combinan ambos parámetros de validez (sensibilidad y especificidad) en la decisión sobre el uso de la prueba evaluada (Fuente-Alba & Molina Villagra, 2017).

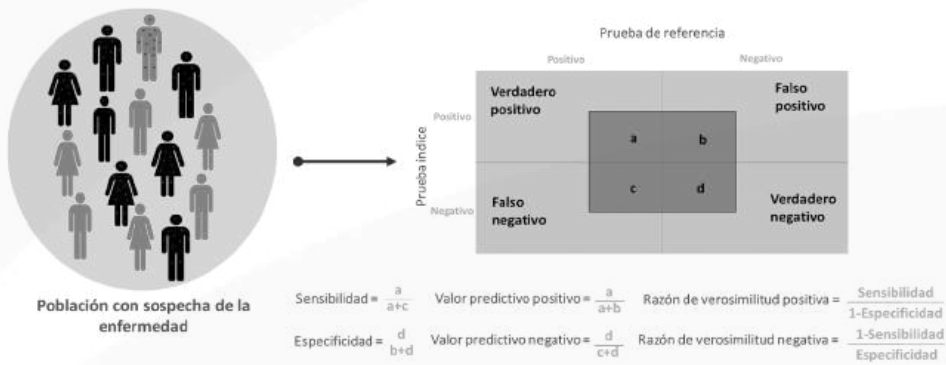
Tabla 1
Valores guía para la interpretación de las razones de verosimilitud

Razones de verosimilitud y su utilidad		
Razón de verosimilitud positiva	Razón de verosimilitud negativa	Utilidad
10	<0,1	Altamente relevante
5 – 10	0,1 - 0,2	Buena
2 – 5	0,2 - 0,5	Regular
< 2	> 0,5	Mala

Nota. Elaboración propia con base en el artículo de Fuente-Alba & Molina Villagra M. (2017). Para la interpretación de los valores de referencia es importante tener presente el evento de interés y la disponibilidad de pruebas para diagnosticarlo. Por ejemplo, en VIH aceptar como excelente 10 VP por cada FP no es muy bueno, allí el CPP debería ser más alto (más exigente). Pero en algunas micosis o en neurocisticercosis, incluso el valor entre 5-10 podría ser excelente por la baja disponibilidad de pruebas o por la complementariedad con datos clínicos o la historia del paciente (familiar, nexo epidemiológico, etc.)

Para calcular la sensibilidad y la especificidad de una prueba es necesario, como se enunció previamente, aplicar una prueba diagnóstica de referencia que clasifique inequívocamente a todos los individuos con o sin la enfermedad; es decir, se debe saber quién “realmente” tiene la enfermedad y quién no la tiene a partir de una fuente externa que no sea la prueba que se está evaluando. Con una prueba diagnóstica ideal, todas las personas se podrían clasificar en dos categorías, los verdaderos positivos que son aquellos con la enfermedad que la prueba índice clasifica correctamente como “positivos” y los verdaderos negativos que son aquellas personas sin la enfermedad que la prueba índice las clasifica correctamente como “negativas”. Sin embargo, rara vez se presenta este escenario. Lo común es que algunas personas que no tienen la enfermedad sean clasificadas de manera errónea como “positivas” por la prueba índice (falsos positivos), y algunas personas con la enfermedad sean clasificadas de manera errónea como “negativas” (falsos negativos) por la prueba índice (ver Figura 1) (Celentano et al., 2024).

Figura 1
Diseño de un estudio de pruebas diagnósticas



Nota. Elaboración propia.

El tema de los falsos positivos es importante porque las personas que dieron un resultado positivo pueden ser sometidas a pruebas más sofisticadas y costosas, o tal vez a un procedimiento invasivo innecesario, generando costos, carga para el sistema de salud y afectaciones emocionales en los pacientes. Por otro lado, si una persona tiene la enfermedad, pero se le informa erróneamente que el resultado de la prueba es negativo, el problema es realmente grave porque podrían constituir, en el peor de los casos, una sentencia de muerte (Celentano et al., 2024).

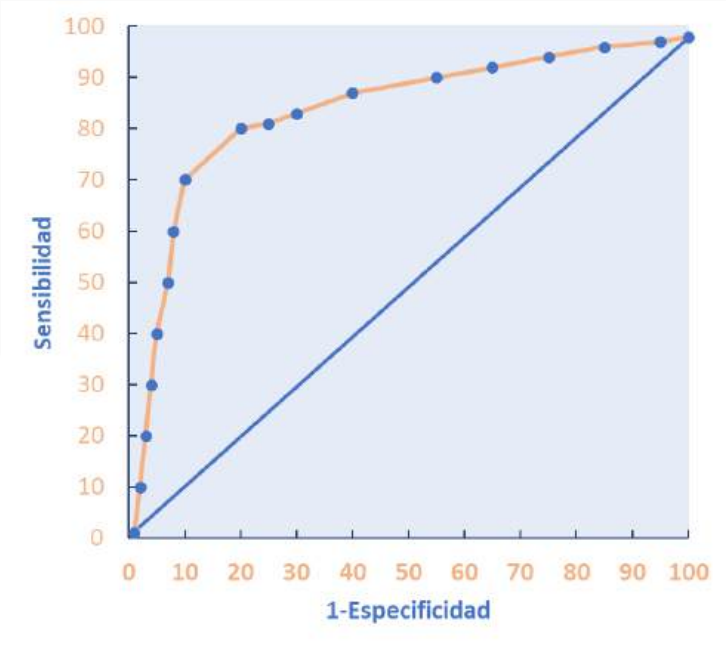
Hasta este punto se han descrito parámetros útiles para pruebas diagnósticas que tienen dos posibles resultados: positivo o negativo. No obstante, **existen muchas pruebas en las que los resultados se presentan en variables cuantitativas (hemoglobina glicosilada, glicemia, presión arterial, etc.) para las cuales es necesario decidir un umbral de positividad** (Celentano et al., 2024). Definir el criterio de positividad plantea varios retos porque tendrá un gran efecto en la sensibilidad y especificidad de una prueba. Por un lado, si se elige un punto de corte muy bajo, la prueba clasificará a todos los individuos como enfermos y tendría una sensibilidad muy alta (cercana al 100 %), pero buena parte de las personas sin la enfermedad también serían clasificadas como enfermas (especificidad cercana a 0 %). Por otro lado, si se selecciona un criterio de positividad muy alto, una gran proporción de pacientes serían clasificados como negativos (especificidad cercana al 100 %), pero no detectaría

la mayoría de los pacientes con la enfermedad (sensibilidad cercana a 0 %).

Por tanto, siempre es importante observar en conjunto la sensibilidad y la especificidad de la prueba. En este sentido, la sensibilidad y especificidad para diferentes puntos de corte en una prueba diagnóstica pueden representarse mediante una curva ROC (ver Figura 2). La gráfica proporciona una representación visual de la relación entre la sensibilidad y la especificidad a medida que cambia el umbral de positividad. El poder discriminatorio de una prueba aumenta cuanto más se acerca la curva a la esquina superior izquierda del gráfico (donde la sensibilidad y la especificidad son del 100 %).

Esta información se puede evaluar de manera cuantitativa calculando el área bajo la curva ROC (AUC) cuyo valor está entre 0 y 1. El AUC para una prueba perfecta es 1.0 y una prueba sin poder discriminatorio entre sanos y enfermos tiene un AUC de 0.5, así que valores de AUC sustancialmente más altos que 0,5 indicarán que la prueba tiene valor diagnóstico. En la Figura 2 se muestra una curva ROC con un AUC cercano a 0,8. La línea diagonal representa una prueba diagnóstica con un AUC de 0,5 que no tiene poder discriminatorio (Bouter et al., 2018; Campbell et al., 2017).

Figura 2
Curva ROC



Nota. Elaboración propia.

Otra forma de evaluar la exactitud de una prueba diagnóstica, combinando la sensibilidad y la especificidad en una medida única, es con una medida epidemiológica: la odds ratio (OR). En el contexto de las pruebas diagnósticas, la OR se denomina como odds ratio diagnóstica (ORD) y se calcula con la odds de tener un resultado positivo si se está enfermo dividido entre la odds de tener un resultado positivo en quienes no tienen la enfermedad $(a \times d) / (c \times b)$ (Lee-flang, 2014). La ORD puede tomar valores entre cero e infinito. Una ORD de 1 indica que la prueba no es discriminante entre personas con y sin la enfermedad, una ORD de menos de 1 significa que un resultado negativo en la prueba es más probable entre las personas con la enfermedad, por lo que la prueba es inútil, y una ORD mayor de 1 significa que un resultado positivo de la prueba es más probable entre las personas con la enfermedad. Es importante precisar que cuanto mayor es la ORD mejor es la prueba, pero desafortunadamente no es posible indicar unos límites (Bouter et al., 2018).

Finalmente, en la evaluación de pruebas diagnósticas es importante considerar la cuestión de la concordancia; es decir que si se repite una prueba, independientemente del observador, se llega al mismo acuerdo⁴. Para dar cuenta de ello se pueden calcular medidas de concordancia, porque, con frecuencia, las opiniones de los profesionales de la salud difieren, no solo de un profesional a otro, sino también en un mismo profesional en diferentes momentos. Estas diferencias se presentan en conductas terapéuticas, en el examen físico y, por supuesto, en las pruebas diagnósticas (Bouter et al., 2018). Esto es importante porque, independientemente de la sensibilidad y la especificidad de una prueba, si no hay concordancia entre los profesionales que la interpretan, la prueba no es útil (Celentano et al., 2024). Para valorar la concordancia de una prueba diagnóstica se calculan generalmente dos parámetros: el Kappa de Cohen, para desenlaces categóricos, y el coeficiente de correlación intraclass para los desenlaces continuos o discretos.

El Kappa de Cohen es una medida del acuerdo entre observadores (patólogo, radiólogo, microbiólogo, bacteriólogo, etc.) que realizan una misma prueba diagnóstica cuando esta tiene resultados categóricos (positivo/negativo, leve/moderada/grave). El índice de Kappa puede tomar valores entre 0 (acuerdo nulo) y 1 (acuerdo perfecto). El acuerdo de 0,80 o más se considera como un acuerdo fuerte entre evaluadores, mientras que el acuerdo de 0,40 o menos

⁴ Es importante remarcar que, en estos casos, los estadísticos de concordancia no son atributos de la prueba diagnóstica, sino de las personas que analizan sus resultados; por ello se enfatiza que son indicadores de consistencia intra (el mismo sujeto en momentos diferentes) o inter-observador (dos observadores independientes).

se considera bajo (ver Tabla 2). No obstante lo anterior, en la práctica clínica es común hallar pruebas diagnósticas con kappas entre 0,40 a 0,70 (Bouter et al., 2018).

Para determinar la concordancia de los resultados de una prueba diagnóstica con un resultado continuo (presión arterial, glicemia, perfil lipídico) por diferentes observadores, se utiliza el coeficiente correlación intraclase (CCI). El CCI también toma valores entre 0 y 1; valores cercanos al cero indican un grado de acuerdo pobre o nulo, y valores cercanos al uno indican un grado de acuerdo perfecto entre observadores (Mandeville, 2005) (ver Tabla 2).

Tabla 2
Valores guía para la interpretación del Kappa y el coeficiente de correlación intraclase (CCI)

Interpretación del índice de Kappa y el coeficiente de correlación intraclase			
Valor de Kappa	Interpretación del grado de acuerdo	Valor del CCI	Interpretación del grado de acuerdo
< 0,20	Ninguno	0	Pobre
0,21 – 0,39	Mínima	0,01 – 0,20	Leve
0,40 – 0,59	Pobre	0,21 – 0,40	Regular
0,60 -0,79	Moderada	0,41 – 0,60	Moderado
0,80 – 0,90	Fuerte	0,61 – 0,80	Sustancial
> 0,90	Casi perfecta	0,81 – 1,00	Casi perfecto

Nota. Elaboración propia.

Para finalizar, en la Tabla 3 se resumen las principales características de los estudios de pruebas diagnósticas.

Tabla 3
Características distintivas de un estudio de evaluación del desempeño de pruebas diagnósticas

Característica	Estudio de Evaluación del Desempeño de Pruebas Diagnósticas
Objetivo principal	Evaluar una prueba diagnóstica en la detección de una enfermedad, comparando sus resultados con una prueba de referencia.
Tipo de estudio	Observacional analítico, generalmente transversal.
Dirección del estudio	Se compara la prueba diagnóstica en cuestión con un criterio diagnóstico de referencia. Los resultados de ambas pruebas se obtienen de manera simultánea.

Selección de participantes	Los participantes son seleccionados de acuerdo con la sospecha de la condición o enfermedad a evaluar.
Muestra	La muestra debe incluir tanto personas con la enfermedad (infección o evento) como personas sin la enfermedad, para poder calcular las medidas de validez y seguridad.
Medición de la exposición	No hay exposición en el sentido convencional; el enfoque está en la aplicación de la prueba diagnóstica, cuyas mediciones se comparan con el estándar de referencia.
Costo	Puede variar, porque depende de la complejidad de las pruebas diagnósticas a comparar.
Tiempo requerido	Corto, ya que se pueden obtener los resultados de la prueba diagnóstica y la prueba de referencia simultáneamente.
Principal medida de asociación	Sensibilidad, especificidad, valor predictivo positivo (VPP), valor predictivo negativo (VPN), razón de verosimilitud (LR), área bajo la curva (AUC) de la curva ROC, concordancia.
Ventajas	Permite evaluar las pruebas diagnósticas; útil para establecer el valor de nuevas pruebas en comparación con las existentes.
Desventajas	Depende de lo invasiva, la calidad, los costos y la disponibilidad de la prueba de referencia; puede haber sesgos si la selección de la muestra no es adecuada.

Nota. Elaboración propia.

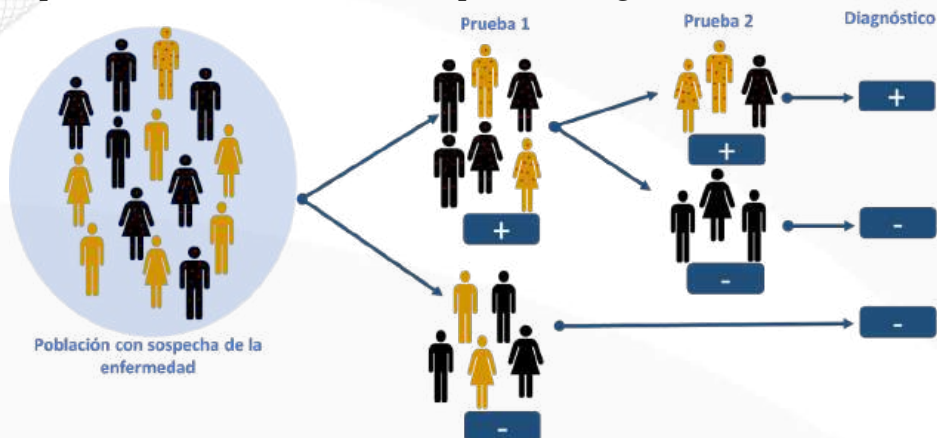
2. Variantes en los estudios de pruebas diagnósticas

En situaciones clínicas en las que existe un alto grado de incertidumbre diagnóstica, el médico podría considerar que una única prueba no es suficiente para iniciar un tratamiento (alcanzar el denominado umbral test-tratamiento), suspender una terapia o declarar al paciente libre de enfermedad (alcanzar el denominado umbral de prueba), de manera que puede considerar enviar pruebas diagnósticas adicionales. En algunas ocasiones esas pruebas diagnósticas son realizadas de manera simultánea y se combinan los resultados para llegar a la decisión diagnóstica; estos estudios se conocen como evaluación de pruebas diagnósticas con diseño en paralelo.

Por otro lado, existen escenarios clínicos que demandan la realización de varias pruebas diagnósticas en secuencia, separadas en el tiempo, de tal manera que los resultados de una prueba se conocen antes de realizar las posteriores. Este diseño se conoce como evaluación de pruebas diagnósticas en serie o pruebas condicionales (ver Figura 3).

Figura 3

Representación de los estudios de pruebas diagnósticas secuenciales



Nota. Elaboración propia.

3. Sobre la validez interna y externa en los estudios de pruebas diagnósticas

Como hemos mencionado, un estudio sobre pruebas diagnósticas en circunstancias ideales incluye pacientes consecutivos o una selección aleatoria de aquellos con sospecha clínica de la enfermedad de interés; así se obtiene un espectro de pacientes representativos de lo que se presenta en la práctica clínica. Luego se aplica la prueba índice (la prueba de diagnóstico que se está evaluando) y la prueba de referencia (la mejor prueba disponible que verifica la presencia o ausencia de una enfermedad) a todos los participantes; finalmente se comparan los resultados de manera independiente. Sin embargo, los investigadores se enfrentan a múltiples limitaciones relacionadas con el presupuesto, la logística o aspectos éticos que impiden que se puedan diseñar estudios en las circunstancias ideales. Por tanto, durante el diseño y la ejecución de estudios sobre pruebas diagnósticas se pueden presentar sesgos que conducen a sobrestimar o subestimar las características de la prueba y, en última instancia, a errores en la práctica clínica (Hall et al., 2019).

Aunque se han identificado numerosos aspectos que pueden afectar la validez interna y externa de las pruebas diagnósticas, en esta sección haremos referencia a los aspectos relacionados con la selección de los participantes, el enmascaramiento en la interpretación de las pruebas, los flujos de tiempo entre las dos pruebas y la

independencia entre la prueba índice y la prueba de referencia.

Durante la selección de los participantes se puede presentar el sesgo de verificación parcial o el sesgo de espectro. En circunstancias ideales, los estudios de evaluación de pruebas diagnósticas deben aplicar la prueba índice y la prueba de referencia de forma independiente. No obstante, en algunos casos se presenta que los investigadores seleccionan a los participantes en función de si tienen o no una prueba de referencia positiva o negativa, lo que se conoce como sesgo de verificación parcial (Hall et al., 2019).

El sesgo de verificación parcial es problemático porque es más probable que los pacientes reciban una prueba de referencia en la práctica clínica cotidiana si antes tienen una prueba índice positiva, así se aumentan artificialmente los verdaderos positivos y la sensibilidad. El sesgo de verificación parcial podría evitarse realizando la prueba de referencia a una muestra aleatoria de personas con sospecha de la enfermedad que resultaron negativas en la prueba índice (Hall et al., 2019).

Por su parte, el sesgo de espectro hace referencia a que las características de los pacientes incluidos en la evaluación de la prueba diagnóstica son sistemáticamente diferentes, en signos, síntomas y gravedad (espectro de la enfermedad), de lo que se presenta en la práctica clínica en la que tendría aplicabilidad la prueba, dando como resultado una población de estudio no representativa, lo que dificultaría la generalización de las características de la prueba de índice (Hall et al., 2019).

El sesgo de espectro se puede presentar por un diseño de casos y controles, la exclusión de pacientes con resultados indeterminados o por un muestreo por conveniencia (Hall et al., 2019). La mayoría de los estudios de pruebas diagnósticas son transversales; no obstante, cuando una enfermedad es poco frecuente, los estudios transversales no son factibles y los investigadores optan por un diseño de casos (pacientes que se sabe tienen la enfermedad) y controles (personas que se sabe están sanas), los cuales se reúnen y se les aplica la prueba índice (Hall et al., 2019). En este contexto, no se incluye el espectro completo de la enfermedad porque los casos representan los más enfermos y los controles los completamente sanos o más saludables. Estudios previos han demostrado que este diseño introduce la mayor sobreestimación de las características de desempeño de una prueba diagnóstica, ya que la OR diagnóstica se sobreestima multiplicándola por un factor de 3 (Lijmer et al., 1999).

El sesgo de espectro también puede presentarse por la exclusión de sujetos con resultados indeterminados o técnicamente difíciles de interpretar. En este sentido, se recomienda que en pruebas que arrojen este tipo de resultados indeterminados, los investigadores declaren explícitamente *a priori* si estos resultados se considerarán positivos o negativos en el análisis. En el sesgo de espectro por muestreo a conveniencia puede haber errores sistemáticos en la selección de pacientes; por ejemplo, los investigadores podrían evitar incluir pacientes en los que la prueba índice es difícil de realizar (similar al de resultados indeterminados), lo que daría como resultado una sensibilidad o especificidad falsamente incrementada, dependiendo de qué pacientes sean excluidos y el estado de enfermedad (Hall et al., 2019).

Con respecto a la independencia en la aplicación de la prueba índice y la de referencia o el enmascaramiento, en la práctica clínica es útil que quien realiza una prueba diagnóstica tenga todos los antecedentes del paciente, resultados de pruebas diagnósticas previas o la impresión clínica de otro profesional de la salud, así puede ser más observador de los hallazgos de su prueba, centrarse en una imagen específica (en el caso de pruebas imagenológicas) o poner más atención en un sitio anatómico particular. En este sentido, la información con la que se cuenta al momento de la interpretación de los resultados de una prueba puede influenciar la interpretación de esa prueba y derivar en un sesgo de revisión.

En el contexto de la evaluación de las pruebas diagnósticas, el sesgo de revisión ocurre cuando aquellos quienes interpretan la prueba índice no están cegados al resultado de la prueba de referencia (Kea et al., 2019), de manera que el operario podría potencialmente alterar su interpretación para estar de acuerdo con los resultados de la prueba de referencia. Este aspecto es más importante cuanto mayor componente de interpretación subjetiva tenga la prueba diagnóstica en cuestión. Lo anterior trae como consecuencia un falso aumento en la sensibilidad y la especificidad de la prueba índice (Kea et al., 2019). De este modo, se ha descrito que la falta de enmascaramiento sobreestima el desempeño global de una prueba en un 30 % (Abraira, 2006; Lijmer et al., 1999).

En relación con la independencia de las pruebas aplicadas, se ha mencionado de manera reiterativa que en los estudios de pruebas diagnósticas los participantes deben recibir tanto la prueba índice como la prueba de referencia. Generalmente, la prueba de referencia es un procedimiento costoso, invasivo o poco ético en quienes

tengan resultados de una prueba índice negativa, por lo que algunos investigadores optan por aplicar una prueba de referencia diferente (menos invasiva, menos costosa) a quienes dieron negativo en la prueba índice. En muchos estudios, particularmente los que se realizan con datos de la práctica clínica cotidiana, la prueba de referencia se realiza a todas las personas que presentaron un resultado positivo en la prueba índice, pero solo a un subgrupo pequeño de las personas que tuvieron un resultado negativo, lo que da lugar al sesgo de verificación parcial (Kea et al., 2019). Otros investigadores tratan de enfrentar esta limitación aplicando a las personas con un resultado negativo en la prueba índice otra prueba de referencia diferente (menos invasiva, menos costosa); esto se conoce como sesgo de verificación diferencial o de doble estándar (Kea et al., 2019). En ambos casos se incumple el criterio de independencia entre las técnicas diagnósticas y ocasionarían una sobreestimación del desempeño diagnóstico. Se ha descrito que cuando se usan diferentes patrones de referencia, el desempeño global de la prueba índice se sobreestima multiplicándose por dos (Lijmer et al., 1999). Adicionalmente, en escenarios clínicos en los que la prueba de referencia es combinada y las pruebas evaluadas hacen parte del estándar, se puede cometer un sesgo que se conoce como sesgo de incorporación (Rendón-Macías et al., 2020), que incrementa falsamente la sensibilidad y la especificidad de la prueba índice (Kea et al., 2019).

Por último, **los resultados de la prueba índice y la prueba de referencia deberían recolectarse de manera simultánea en los mismos pacientes y en el mismo periodo de tiempo**, ya que si hay un retraso entre pruebas la enfermedad puede progresar, el paciente puede iniciar tratamiento o incluso recuperarse entre el periodo de tiempo que se tarda entre la aplicación de la prueba de referencia y la prueba índice, generando sesgos por errores de clasificación de los pacientes. En este caso es importante tener en cuenta que el riesgo de este tipo de sesgos varía dependiendo de la enfermedad a estudiar: en enfermedades infecciosas agudas, no aplicar las pruebas de manera simultánea podría invalidar los resultados, pero en enfermedades infecciosas crónicas el retraso de pocos días en la aplicación de una de las pruebas no debería considerarse un problema (Whiting et al., 2011).

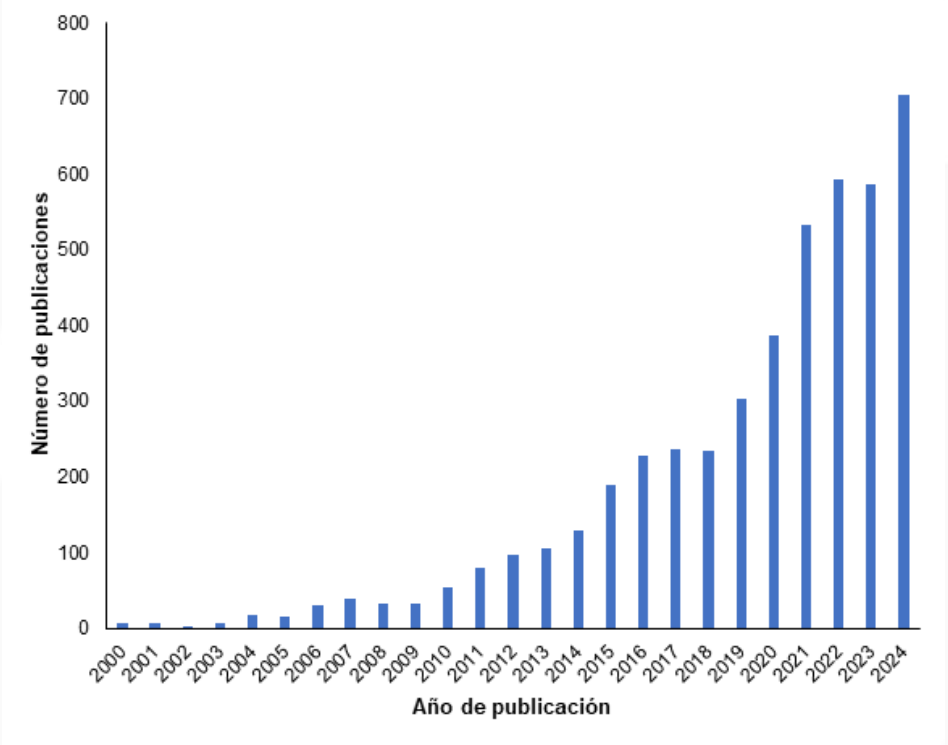
4. Indicaciones para sistematizar la evidencia de los estudios sobre pruebas diagnósticas y elegir cuándo combinar los resultados

Los grandes avances tecnológicos y científicos han permitido el desarrollo continuo de más y mejores pruebas diagnósticas que llegan a competir con las existentes por ser más económicas, más rápidas, más fáciles de realizar, menos invasivas y más precisas. En consecuencia, es posible que se encuentren múltiples tecnologías disponibles, con características de precisión muy heterogéneas, para el diagnóstico de una misma enfermedad, lo que constituye un reto para los profesionales de la salud porque deben argumentar sus decisiones sobre usar o no una prueba diagnóstica en sus pacientes con base en una copiosa evidencia científica. En este contexto, emergen las revisiones sistemáticas de la precisión de pruebas diagnósticas porque proporcionan un resumen basado en toda la evidencia disponible de las alternativas diagnósticas para una enfermedad, evalúan la calidad de los estudios publicados y explican la variación en los hallazgos entre esos estudios (Campbell et al., 2017).

Las revisiones sistemáticas y metanálisis en el campo de la evaluación de pruebas diagnósticas se realizan principalmente en cinco escenarios: i) cuando se desean obtener estimaciones resumidas más válidas y generalizables de la precisión de una prueba diagnóstica; ii) cuando se quiere sistematizar todas las pruebas diagnósticas disponibles para una misma enfermedad; iii) en los casos en los que hay interés en conocer si una nueva prueba diagnóstica es mejor que las alternativas actuales, pero los estudios individuales no son concluyentes o tienen discrepancias entre ellos; iv) cuando se busca identificar los factores (covariables) que afectan la precisión de la prueba diagnóstica en diferentes estudios individuales; y v) para perfilar o identificar vacíos en el conocimiento sobre los cuales hay que desarrollar investigaciones futuras (Kim et al., 2015).

En este sentido, se han realizado una amplia diversidad de revisiones sistemáticas de estudios sobre pruebas diagnósticas. Una búsqueda en PubMed, con los términos (Metanalysis[Title]) AND (diagnosis [Title/Abstract]) AND (Accuracy [Title/Abstract]) arroja 3822 resultados, mientras que con la búsqueda (systematic review[Title]) AND (diagnosis [Title/Abstract]) AND (Accuracy [Title/Abstract]) se encuentran 4647 publicaciones relacionadas con la síntesis de la literatura de estudios de pruebas diagnósticas entre el 2000 y el 2024 (ver Figura 4). Los estudios se han realizado sobre pruebas diagnósticas de laboratorio, imagenológicas y de algunas escalas.

Figura 4
Frecuencia de publicaciones en PubMed sobre revisiones sistemáticas de estudios de pruebas diagnósticas entre los años 2000 y 2024



Nota. Elaboración propia.

Entre las pruebas diagnósticas de laboratorio se encuentran estudios sobre la precisión diagnóstica de las pruebas de esterasa leucocitaria y nitrito para el diagnóstico de bacteriuria en adultos mayores (Moragas et al., 2026), de las pruebas rápidas POCT para el diagnóstico de infecciones respiratorias agudas (Webster et al., 2024), biomarcadores séricos para la enterocolitis necrosante (Pimenta et al., 2025), entre otros. Entre los estudios sobre pruebas imagenológicas se han evaluado las imágenes diagnósticas de ecografía para el cáncer de tiroides (Liu & Xu, 2025), de técnicas de imagen para la enfermedad de Alzheimer (Radmard et al., 2025), tomografía por emisión de positrones, resonancia magnética funcional y las imágenes por tensor de difusión el diagnóstico de afecciones neurodegenerativas (Chandrasekar et al., 2025). Este método también se ha aplicado para probar la precisión diagnóstica de escalas; algunos ejemplos son las escalas el diagnóstico de deshidratación (Falszewska et al., 2018), de ansiedad y depresión (Brennan et al., 2010) o del déficit de atención e hiperactividad (Chang et al., 2016).

4.1 Características del protocolo de búsqueda y selección de estudios en la síntesis de investigaciones sobre pruebas diagnósticas

Aunque todas las revisiones sistemáticas parten de una pregunta de investigación que tiene unos principios similares, la revisión sistemática de pruebas diagnósticas requiere de una pregunta con características particulares que, si bien puede seguir la estrategia PICO, cada letra tiene una connotación diferente. La “P” se refiere a la población con la enfermedad de interés y algunas características de los pacientes que pueden alterar las características de desempeño de la prueba. La “I” hace alusión a la prueba índice, la prueba objeto de la evaluación. La “C” será la prueba de referencia, aquella que clasifica con certeza los verdaderos positivos y los verdaderos negativos. Finalmente, en la “O” se expresan los parámetros de evaluación de la prueba que se quieren comparar (sensibilidad, especificidad, valores predictivos, razones de verosimilitud, curva ROC, OR diagnóstica) (Kim et al., 2015).

Otros autores sugieren que en lugar de estrategia PICO podría ser nombrada como estrategia PIRD (Población, Índice, Referencia, Diagnóstico) (Campbell et al., 2017).

4.1.1 La estrategia de búsqueda

La descripción de la estrategia de búsqueda es un elemento clave para la reproducibilidad y la exhaustividad de una revisión sistemática. Permite a los lectores evaluar las decisiones tomadas por los investigadores para identificar los artículos de interés. La estrategia de búsqueda debe ser coherente con la pregunta PICO (PIRD), y las palabras clave elegidas deben especificarse junto con las bases de datos a buscar. Se recomienda que la estrategia de búsqueda se desarrolle en tres etapas: i) una búsqueda preliminar en las bases de datos seleccionadas, utilizando algunas palabras clave relacionadas con la enfermedad de interés (y sus homólogas), la prueba índice, la prueba de referencia y parámetros de evaluación diagnóstica; esto se hace con el propósito de identificar términos adicionales mediante el análisis de conceptos contenidos en el título, palabras clave, resumen y lista de indexación. ii) Una búsqueda específica en las bases de datos para identificar palabras clave relevantes que no fueron tenidas en consideración en la búsqueda inicial. Y iii) una revisión

de las referencias de los estudios incluidos con el propósito de identificar estudios no captados en los dos pasos anteriores (Campbell et al., 2017).

Además, algunos autores complementan esta estrategia consultando (vía email, por ejemplo) a investigadores expertos en el campo de interés quienes pueden ser una fuente potencial de artículos o datos no publicados (Campbell et al., 2017). Una estrategia de búsqueda básica para hallar artículos sobre pruebas diagnósticas es enumerar las palabras homólogas para cada elemento de la pregunta PICO y combinarlas usando “OR” dentro de la columna y “AND” entre las columnas, como se muestra en la Tabla 4.

Tabla 4
Estrategia de búsqueda básica para una revisión sistemática de pruebas diagnósticas

Ejemplo de una estrategia de búsqueda básica						Base de datos		
Población (P)	Prueba índice (I)		Prueba de referencia (C)		Parámetro diagnóstico (O)	Pubmed	Scielo	Google scholar
Liver cancer	AND	Magnetic resonance imaging	AND	Computed tomography	Diagnosis	n	n	n
						n	n	n
						n	n	n
						n	n	n
Hepatocellular cancer					Sensitivity	n	n	n
						n	n	n
						n	n	n
						n	n	n
Liver carcinoma					Specificity	n	n	n
						n	n	n
						n	n	n
						n	n	n
Hepatocellular carcinoma					ROC curve	n	n	n
						n	n	n
						n	n	n
						n	n	n

Nota. Elaboración propia.

En este punto es importante precisar que algunas bases de datos han desarrollado filtros metodológicos basados en términos de indexación con la intención de arrojar resultados más específicos cuando se agregan estos filtros. Sin embargo, un estudio previo demostró que los filtros de búsqueda no funcionan de manera consistente y no deberían usarse como el único enfoque en las búsquedas para realizar revisiones sistemáticas de esta modalidad de investigación (Beynon et al., 2013).

4.1.2 Los criterios de inclusión

El factor más importante para la definición de los criterios de inclusión es la relevancia de los artículos identificados para responder a la pregunta de investigación; así que los criterios de inclusión generalmente se expresan en el formato de pregunta PICO (PIRD) (Campbell et al., 2017; Kim et al., 2015).

En la población se debe garantizar que los tipos de participantes sean apropiados para los objetivos de la revisión y reflejar quién se someterá a la prueba diagnóstica en la práctica clínica. De esta manera, en las razones para la inclusión de los artículos se pueden enunciar aspectos relacionados con el estadio de la enfermedad, los síntomas y otro tipo de características de los pacientes como la edad, el sexo, entre otros. La prueba índice es la prueba diagnóstica cuya precisión se está investigando en la revisión y su nombre debe quedar explícito en los criterios de inclusión; sin embargo, en ocasiones pueden existir múltiples pruebas con principios diagnósticos similares, por lo que se debe especificar en el protocolo qué criterios se utilizarán para determinar si las pruebas son lo suficientemente similares para incluirlas en la revisión (Campbell et al., 2017).

La prueba de referencia es aquella frente a la cual se compararán los resultados de la prueba de índice. Esta debería ser la mejor prueba disponible para el diagnóstico de la enfermedad de interés, y su nombre debe quedar explícito en los criterios de inclusión. Además, al igual que para la prueba índice, pueden existir varias pruebas de referencia que compartan los mismos estándares de calidad, de modo que se deben especificar los criterios que se utilizarán para determinar las pruebas que se considerarán como prueba de referencia (Campbell et al., 2017).

Finalmente, se definen los tipos de estudio que podrán incluirse en la revisión. Usualmente, los tipos de estudio serán aquellos que tengan un diseño de precisión de pruebas diagnósticas; no obstante, como los estudios de pruebas diagnósticas usan diseños transversales, los datos de interés para la revisión pueden recuperarse de estudios cuyo propósito no era necesariamente evaluar la precisión de la prueba. Sin embargo, es importante tener en cuenta que en enfoque amplio no sería necesario ser tan estrictos en la búsqueda de la prueba de referencia (esta puede ser una variable por analizar, o una variable para hacer meta-regresiones).

4.1.3 Los criterios de exclusión

Los criterios de exclusión se formulan con base en factores que puedan generar sesgos debido a un diseño de estudio inapropiado o de baja calidad. En este sentido, los criterios de exclusión contienen aspectos como: i) definición imprecisa y ambigua de la enfermedad de interés o los criterios de positividad, ii) la población no representa todo el espectro clínico de aquellos a quienes se aplicará la prueba en la práctica clínica, iii) la prueba índice no se aplica a todos los pacientes, iv) el estudio utiliza varias pruebas de referencia y/o esta solo se aplica a una fracción de la población, v) no se garantizaron las condiciones preanalíticas para la realización de la pruebas, vi) la prueba de referencia se aplicó después de conocer el resultado de la prueba índice, y vii) el estudio no reporta los datos de precisión diagnóstica en la tabla de 2 por 2, por lo que no es posible calcular todos los parámetros enunciados en capítulos anteriores.

Es importante tener en cuenta que es posible ser muy restrictivo con la definición de los criterios de exclusión, de manera que se filtran solo investigaciones con altos estándares de calidad; no obstante, existe consenso en preferir el análisis de toda la evidencia disponible flexibilizando los criterios de exclusión y luego comunicar la calidad de las investigaciones halladas, haciendo análisis por subgrupos entre los estudios con diferentes grados de calidad (Ciapponi, 2015) o meta-regresiones.

4.1.4 La extracción de la información

Una vez que se han seleccionado los estudios relevantes para la revisión sistemática y/o el metanálisis, dos revisores deberán extraer la información requerida de manera independiente. Los desacuerdos entre ellos deben resolverse por consenso o por referencia a una tercera persona. En esta etapa, es útil calcular un índice de concordancia entre los revisores (índice Kappa o coeficiente de correlación intraclase) y exponerlo en la metodología de la revisión para dar mayor confianza a los lectores.

Sumado a lo anterior, es recomendable utilizar un formulario estructurado en Microsoft Excel para extraer los datos. El formulario debe incluir como mínimo la siguiente información: i) características del estudio, tales como autores, año de publicación, país de realización, años de reclutamiento de pacientes, tamaño de muestra

utilizado; ii) características demográficas de los pacientes del estudio como su edad y sexo, así como aspectos clínicos relacionados con la enfermedad de interés, los criterios diagnósticos iniciales y comorbilidades en los casos que aplique; iii) características de la prueba índice y la prueba de referencia, el tipo de técnica, el protocolo de la prueba y el criterio de positividad; y iv) los parámetros para estimar la precisión de la prueba índice que pueden extraerse de la tabla de 2 por 2 e incluir el número de verdaderos positivos (VP), falsos positivos (FP), falsos negativos (FN) y verdaderos negativos (VN) (Kim et al., 2015).

4.2 Evaluación de la calidad metodológica y riesgos de sesgos

La guía *Quality Assessment of Diagnostic Accuracy Studies* (QUADAS) ha sido ampliamente utilizada por la colaboración Cochrane y el Instituto Nacional de Excelencia Clínica del Reino Unido (NICE) para evaluar la calidad metodológica de los estudios de precisión diagnóstica. Esta guía se desarrolló por primera vez en el 2003 como una herramienta de evaluación de la calidad de los estudios sobre precisión diagnóstica que se incluirían en las revisiones sistemáticas (Ciapponi, 2015). En la actualidad, existe una versión mejorada denominada QUADAS-2.

La guía QUADAS-2 evalúa un estudio en 4 dimensiones: la selección de los participantes, la prueba a evaluar o índice, la prueba estándar o de referencia y el flujo de los participantes a través del estudio, los momentos en los que se realiza la prueba índice y la prueba de referencia (flujo y tiempos). Cada dimensión se valora teniendo en cuenta el riesgo de sesgo y la aplicabilidad, y se clasifica el riesgo de sesgo en bajo, alto o incierto. Como resultado de la evaluación de la calidad metodológica, las revisiones sistemáticas de esta modalidad de investigación deben incluir los resultados de la evaluación QUADAS-2 para cada estudio incluido en la revisión (ver Tabla 5) y para la totalidad de los estudios (ver Figura 5). Se pueden encontrar más detalles sobre QUADAS-2 en el sitio web oficial (www.Bris.ac.uk/quadas/), la descripción del procedimiento para aplicar la herramienta y ejemplos de evaluaciones QUADAS-2 completas (Ciapponi, 2015).

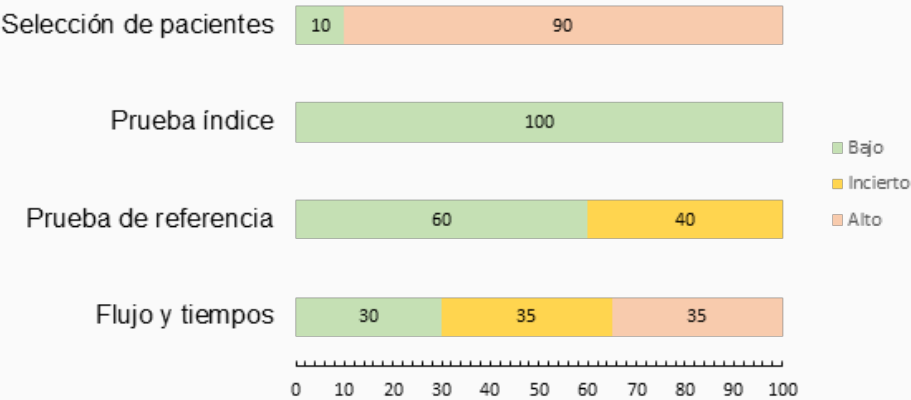
Tabla 5
Representación tabular de la evaluación de la calidad metodológica con QUADAS-2 de los estudios individuales incluidos en la revisión

Estudio	Probabilidad de sesgos				Aplicabilidad de los resultados		
	Selección de pacientes	Prueba índice	Prueba de referencia	Flujo y tiempos	Selección de pacientes	Prueba índice	Prueba de referencia
Estudio 1	😊	😞	😞	?	😞	😞	😞
Estudio 2	😊	😊	😊	😊	😊	😊	😊
Estudio 3	😊	😞	😊	😊	😊	?	😊
Estudio 4	😞	😊	😊	😊	😊	😊	😊
Estudio 5	😊	😊	?	😊	?	?	😞
Estudio 6	😞	😊	😞	😞	😊	😊	😞

😊 Bajo riesgo 😞 Alto riesgo ? Incierto

Nota. Elaboración propia.

Figura 5
QUADAS-2 proporción de estudios incluidos en la revisión con riesgo de sesgo bajo, incierto o alto



Nota. Elaboración propia.

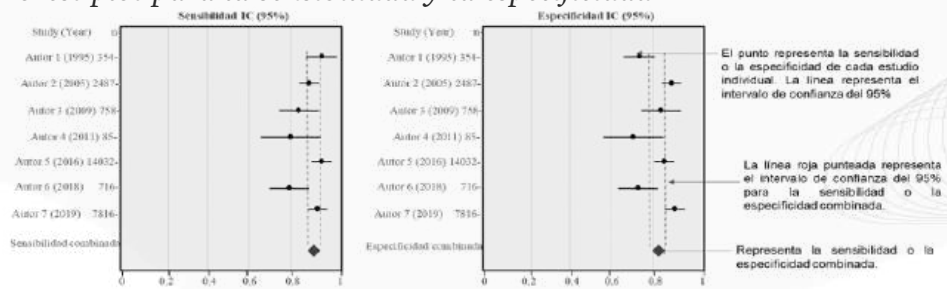
4.3 Análisis de la información

El análisis de la información debe comenzar con un resumen del proceso llevado a cabo desde la búsqueda hasta la selección final de los estudios, mencionando cuántos artículos se han incluido o excluido en cada etapa y las razones para ello. Esta descripción debe ir acompañada de un diagrama de flujo conforme a la declaración PRISMA (PRISMA, 2025). Posteriormente, se debe realizar una descripción general de los estudios incluidos con el fin de dar informa-

ción para que los lectores evalúen cuán similares son los estudios entre sí, incluir las características de los pacientes, los contextos en los que se realizaron las pruebas, una breve descripción de las pruebas índices, la prueba de referencia y el diseño de los estudios. Las tablas con el cálculo de frecuencias absolutas y relativas, junto con intervalos de confianza puede ser la forma más apropiada para presentar estos datos. También se debe incluir un resumen narrativo de la calidad metodológica de los estudios junto con la tabla y los gráficos enunciados en la sección anterior (Campbell et al., 2017).

La síntesis cuantitativa de los estudios sobre pruebas diagnósticas es diferente de la síntesis cualitativa de las demás modalidades de investigación, porque los estudios sobre pruebas diagnósticas deben tratar simultáneamente con varias medidas de resultado (sensibilidad, especificidad, valores predictivos, razones de verosimilitud, curva ROC), mientras que los otros metanálisis usualmente tratan con una sola medida de resultado. En este sentido, se debe incluir el *forest plot* para la sensibilidad, la especificidad, los valores predictivos y las razones de verosimilitud (ver Figura 6). Estos gráficos muestran el parámetro de precisión diagnóstica en cada uno de los estudios individuales incluidos, su intervalo de confianza y la medida combinada. Los resultados más representativos de los gráficos se deben narrar en la prosa, además de informar el número de verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos (Campbell et al., 2017).

Figura 6
Forest plot para la sensibilidad y la especificidad

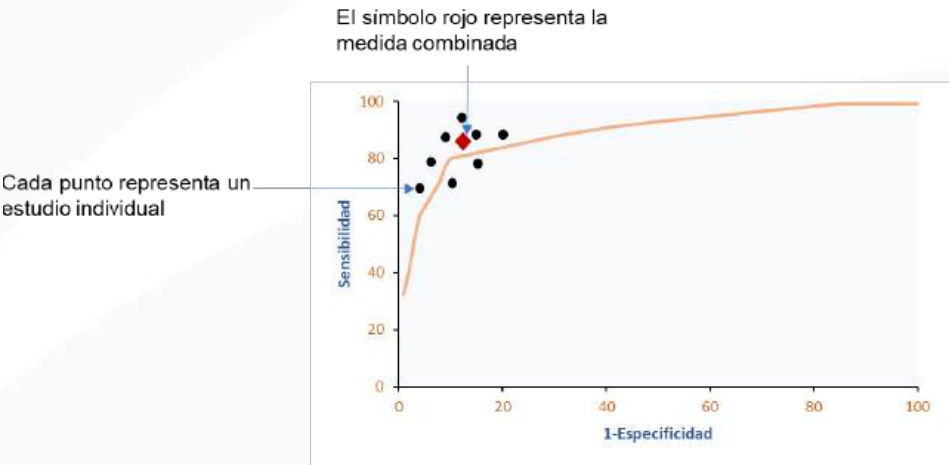


Nota. Elaboración propia.

También, es posible crear curvas ROC en los que cada estudio primario contribuye a un punto único definido por su sensibilidad y especificidad para un umbral específico. El tamaño del punto puede variar según el tamaño de la muestra. Si se ha utilizado el mismo umbral de positividad en cada uno de los estudios primarios, es

apropiado calcular la sensibilidad y especificidad combinada como se muestra en la Figura 7.

Figura 7
Curva ROC para cada estudio individual y la medida combinada



Nota. Elaboración propia.

La mayoría de los programas para este tipo de metanálisis no desagregan resultados de heterogeneidad, sensibilidad y sesgo de publicación. Lo más común es presentar un *forest plot* que incluye los estadísticos de heterogeneidad y sesgo de publicación en la parte inferior, con tablas que indican el porcentaje de peso de cada estudio en la medida combinada, pero sin presentar la medida de resumen que se obtiene al excluir cada uno de los estudios en fases sucesivas (como se hace en los análisis de sensibilidad de otros estudios).

También es común que todas las estimaciones globales se realicen con medidas de efectos aleatorios, dado que para un mismo grupo de datos es posible hallar homogeneidad para un parámetro y heterogeneidad para otros. En términos pragmáticos, si un parámetro de validez diagnóstica registra homogeneidad y se estima el metanálisis con efectos aleatorios, esta adición de la varianza entre estudios no haría variar de manera considerable el intervalo de confianza que se obtendría con una estimación por efectos fijos.

Igualmente, es común que solo se meta analicen los parámetros de validez diagnóstica y no se incluyan los valores predictivos, dado que estos tendrían que ajustarse por la prevalencia del evento en los contextos en que se realizó cada estudio incluido en la revisión sistemática, lo que complejizaría el proceso y no permitiría concentrar

la evaluación en las propiedades intrínsecas que las prueba (las que no cambian por la magnitud del evento).

5. Ejemplo práctico

Problema de investigación: La paracoccidioidomicosis es una micosis sistémica que afecta pulmones, mucosas, sistema nervioso, entre otros órganos. Se han diseñado diferentes pruebas inmunológicas para establecer diagnóstico oportuno, evaluar respuesta al tratamiento y evolución de la infección; sin embargo, los parámetros de evaluación diagnóstica que reportan los estudios individuales son muy variables, con datos de sensibilidad que van desde el 53 % hasta el 100 % y datos de especificidad que oscilan entre 79 % y el 99 %.

Objetivo: evaluar la validez diagnóstica de las pruebas inmunológicas en la infección por *Paracoccidioides* a partir de un metanálisis.

Método: Se utilizó la estrategia de búsqueda: ((*Paracoccidioides* [Title/Abstract]) AND sensitivity [Title/Abstract]) reemplazando en cada búsqueda la palabra sensitivity con specificity, diagnosis, validity. La búsqueda se hizo en las bases de datos ScienceDirect, PubMed, Scopus y Scielo con la aplicación de criterios de inclusión y exclusión.

Resultados: Se identificaron 13 artículos a los cuales se les extrajo la información del número de verdaderos positivos, falsos positivos, falsos negativos y verdaderos negativos, tal como se muestra en la Tabla 6.

Tabla 6
Datos para extraer de cada estudio individual

Autor	Verdaderos positivos (a)	Falsos positivos (b)	Falsos negativos (c)	Verdaderos negativos (d)
Autor 1	158	41	2	159
Autor 2	103	1	11	139
Autor 3	69	13	0	193
Autor 4	22	1	1	92
Autor 5	211	1	7	177
Autor 6	176	2	0	164
Autor 7	68	0	2	49
Autor 8	42	1	4	86

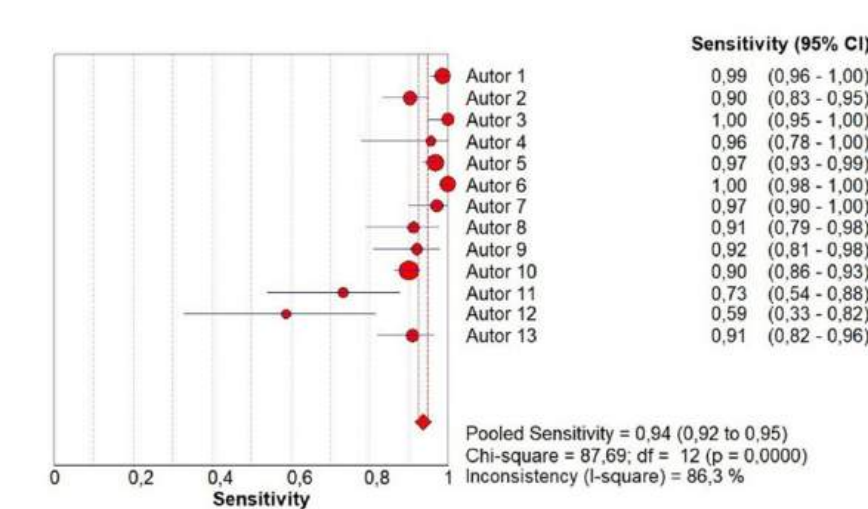
Autor 9	47	1	4	95
Autor 10	316	0	35	200
Autor 11	22	15	8	56
Autor 12	10	1	7	60
Autor 13	70	3	7	63

Nota. Elaboración propia.

La información de esta tabla se introdujo en el software Metadisc (<https://www.itycis.org/es/innovacion/metadisc/>) y se realizó el metanálisis para los parámetros de sensibilidad (Figura 8), especificidad (Figura 9), la razón de verosimilitud positiva (Figura 10) y negativa (Figura 11), la odds ratio diagnóstica (Figura 12) y la curva ROC (Figura 13).

La sensibilidad en los estudios individuales estuvo entre 73 % (IC 95 % 54-88) y 100 % (IC 95 % 98-100). Al combinar los 13 estudios individuales se halló una sensibilidad de 94 % (IC 95 % 92-95) (ver Figura 8).

Figura 8
Forest plot para la sensibilidad

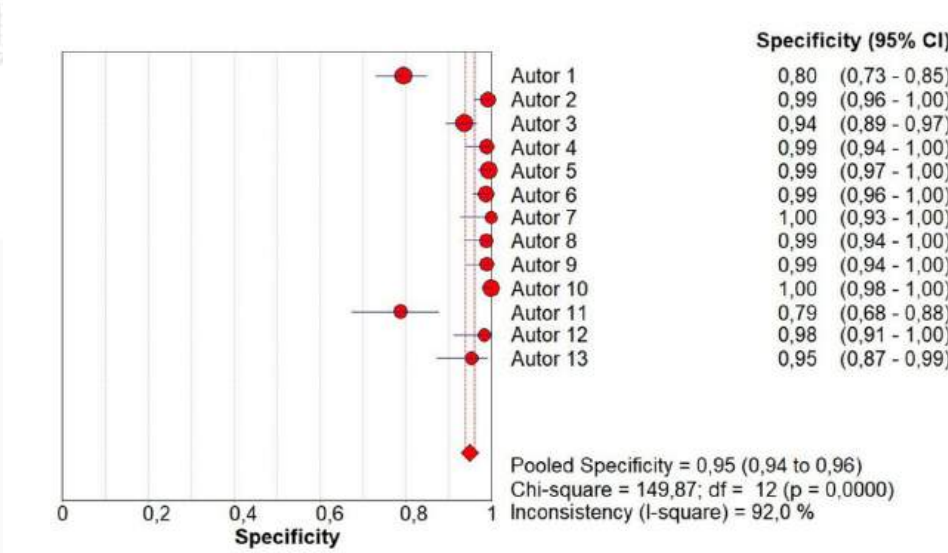


Nota. Elaboración propia.

La especificidad en los estudios individuales estuvo entre 79 % (IC 95 % 68-88) y 100 % (IC 95 % 98-100). Al combinar los 13 estudios individuales se halló una especificidad de 95 % (IC 95 % 94-96) (ver Figura 9).

Figura 9

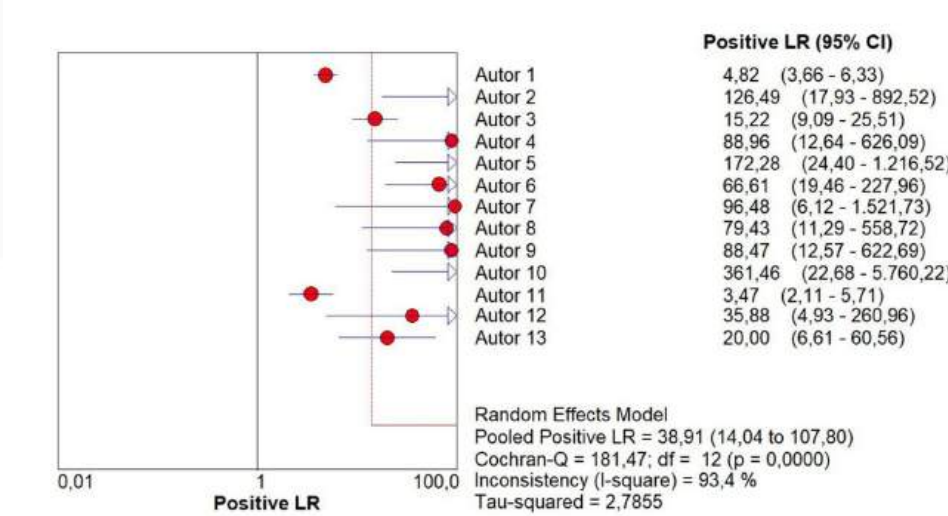
Forest plot para la especificidad



Nota. Elaboración propia.

La razón de verosimilitud positiva en los estudios individuales estuvo entre 3,47 (IC 95 % 2,11-5,71) y 361,4 (IC 95 % 22,68-5760). Al combinar los 13 estudios individuales se halló una razón de verosimilitud positiva de 38,9 (IC 95 % 14-107) (ver Figura 10).

Figura 10
Forest plot para las razones de verosimilitud positiva

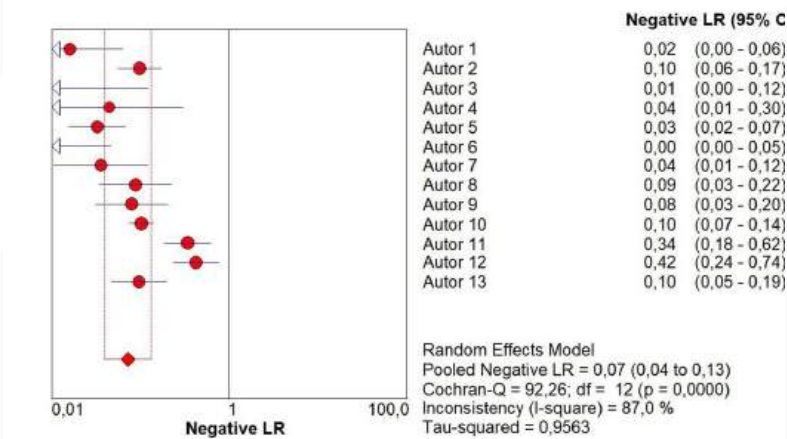


Nota. Elaboración propia.

La razón de verosimilitud negativa en los estudios individua-

les estuvo entre 0,00 (IC 95 % 0,00-0,05) y 0,42 (IC 95 % 0,24-0,74). Al combinar los 13 estudios individuales se halló una razón de verosimilitud negativa de 0,07 (IC 95 % 0,04-0,13) (ver Figura 11).

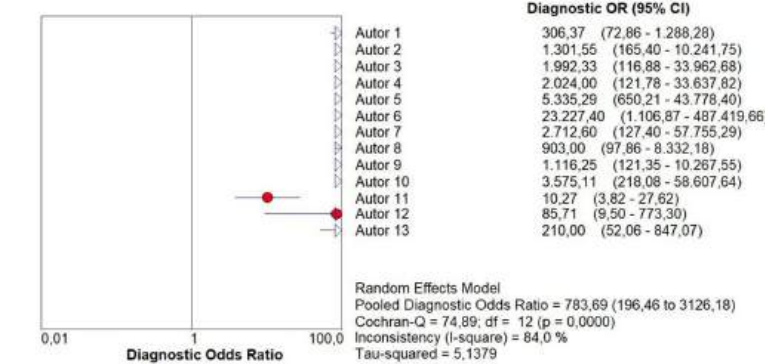
Figura 11
Forest plot para las razones de verosimilitud negativa



Nota. Elaboración propia.

La odds ratio diagnóstica en los estudios individuales estuvo entre 10,27 (IC 95 % 3,82-27,6) y 23227 (IC 95 % 650,2-43778). Al combinar los 13 estudios individuales se halló una odds ratio diagnóstica de 783,6 (IC 95 % 196,5-3126,2) (ver Figura 12).

Figura 12
Forest plot para las OR diagnóstica

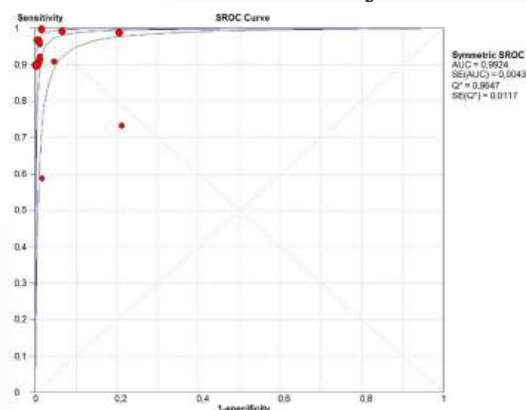


Nota. Elaboración propia.

Finalmente, se muestra que el área bajo la curva ROC fue de 0,99 (ver Figura 13).

Figura 13

Curva ROC para cada estudio individual y la medida combinada



Nota. Elaboración propia.

Conclusión: la heterogeneidad en los estudios individuales es alta; no obstante, las medidas combinadas muestran que las pruebas inmunológicas para el diagnóstico de la infección por *Paracoccidioides* tienen alta validez diagnóstica.

Referencias

- Abraira V. (2006). Sesgos en los estudios sobre pruebas diagnósticas. *SEMERGEN*, 32(1), 24-26. [https://doi.org/10.1016/S1138-3593\(06\)73213-0](https://doi.org/10.1016/S1138-3593(06)73213-0)
- Aschengrau, A., & Seage G.R. (2020). *Essentials of Epidemiology in Public Health* [Fundamentos de epidemiología en salud pública]. Jones & Bartlett Learning
- Beynon, R., Leeflang, M. M., McDonald, S., Eisinga, A., Mitchell, R. L., Whiting, P., & Glanville, J. M. (2013). Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE [Estrategias de búsqueda para identificar estudios de precisión diagnóstica en MEDLINE y EMBASE]. *The Cochrane Database of Systematic Reviews*, (9), MR000022. <https://doi.org/10.1002/14651858.MR000022.pub3>
- Bouter, L. M., Zielhuis, G. A., & Zeegers M. P. A. (2018). *Textbook of Epidemiology* [Libro de texto de epidemiología]. Bohn, Stafleu van Loghum.

- Brennan, C., Worrall-Davies, A., McMillan, D., Gilbody, S., & House, A. (2010). The Hospital Anxiety and Depression Scale: a diagnostic meta-analysis of case-finding ability [Escala de ansiedad y depresión hospitalaria: un metanálisis diagnóstico de la capacidad de detección de casos]. *Journal of Psychosomatic Research*, 69(4), 371–378. <https://doi.org/10.1016/j.jpsychores.2010.04.006>
- Campbell, J. M., Kulgar, M., Ding, S., Carmody, D. P., Hakonsen, S. J., Jadotte, Y. T., White, S., & Munn, Z. (2017). Chapter 9: Diagnostic test accuracy systematic reviews [Capítulo 9: Revisiones sistemáticas sobre la precisión de las pruebas diagnósticas]. In E. Aromataris, & Z. Munn (Eds.), *Joanna Briggs Institute Reviewer's Manual*. The Joanna Briggs Institute. <https://reviewersmanual.joannabriggs.org/>
- Celentano, D. D., Szklo, M., & Farag, Y. M. K. (2024). Gordis. *Epidemiología* (7 Ed.). Elsevier.
- Chandrasekar, S. K., Arthanari, J., Chandrasekar, K. K., Gaviria, E., Shashank, S., Jethi, A., John, J., Bopparaju, S., Potulapati, I., Maateen, M. A., & Hussin, O. A. (2025). Advanced Imaging Techniques (PET, fMRI, DTI) in Early Detection of Neurodegenerative Diseases: A Systematic Review [Técnicas avanzadas de imagen (PET, fMRI, DTI) en la detección temprana de enfermedades neurodegenerativas: una revisión sistemática]. *Health Science Reports*, 8(7), Article e70855. <https://doi.org/10.1002/hsr2.70855>
- Chang, L. Y., Wang, M. Y., & Tsai, P. S. (2016). Diagnostic Accuracy of Rating Scales for Attention-Deficit/Hyperactivity Disorder: A Meta-analysis [Precisión diagnóstica de las escalas de calificación para el trastorno por déficit de atención e hiperactividad: un metanálisis]. *Pediatrics*, 137(3), Article e20152749. <https://doi.org/10.1542/peds.2015-2749>
- Ciapponi A. (2015). QUADAS-2: instrumento para la evaluación de la calidad de estudios de precisión diagnóstica. *Evidencia*, 18(1), 22-26. https://www.researchgate.net/publication/351573778_QUADAS-2_instrumento_para_la_evaluacion_de_la_calidad_de_estudios_de_precision_diagnostica
- Falszewska, A., Szajewska, H., & Dziechciarz, P. (2018). Diagnostic accuracy of three clinical dehydration scales: a systematic review [Precisión diagnóstica de tres escalas de deshidratación clínica: una revisión sistemática]. *Archives of Disease in Childhood*, 103(4), 383-388. <https://doi.org/10.1136/archdischild-2017-313762>

- Fuente-Alba, C. S., & Molina Villagra, M. (2017). Likelihood ratio (razón de verosimilitud): definición y aplicación en Radiología. *Revista Argentina de Radiología*, 81(3), 204-208. <https://doi.org/10.1016/j.rard.2016.11.002>
- Hall, M. K., Kea, B., & Wang, R. (2019). Recognising Bias in Studies of Diagnostic Tests Part 1: Patient Selection [Reconocimiento del sesgo en los estudios de pruebas diagnósticas Parte 1: Selección de pacientes]. *Emergency Medicine Journal*, 36(7), 431-434. <https://doi.org/10.1136/emermed-2019-208446>
- Kea, B., Hall, M. K., & Wang, R. (2019). Recognising bias in studies of diagnostic tests part 2: interpreting and verifying the index test [Reconocimiento del sesgo en los estudios de pruebas diagnósticas parte 2: interpretación y verificación de la prueba índice]. *Emergency Medicine Journal*, 36(8), 501-505. <https://doi.org/10.1136/emermed-2019-208447>
- Kim, K. W., Lee, J., Choi, S. H., Huh, J., & Park, S. H. (2015). Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part I. General Guidance and Tips [Revisión sistemática y metanálisis de estudios que evalúan la precisión de las pruebas diagnósticas: una revisión práctica para investigadores clínicos - Parte I. Orientación general y consejos]. *Korean Journal of Radiology*, 16(6), 1175-1187. <https://doi.org/10.3348/kjr.2015.16.6.1175>
- Leeftang, M. M. (2014). Systematic reviews and meta-analyses of diagnostic test accuracy [Revisiones sistemáticas y metanálisis de la precisión de las pruebas diagnósticas]. *Clinical Microbiology and Infection*, 20(2), 105-113. <https://doi.org/10.1111/1469-0691.12474>
- Lijmer, J. G., Mol, B. W., Heisterkamp, S., Bossel, G. J., Prins, M. H., van der Meulen, J. H., & Bossuyt, P. M. (1999). Empirical evidence of design-related bias in studies of diagnostic tests [Evidencia empírica de sesgo relacionado con el diseño en estudios de pruebas diagnósticas]. *JAMA*, 282(11), 1061-1066. <https://doi.org/10.1001/jama.282.11.1061>
- Liu, Y., & Xu, W. (2025). Application and limitations of ultrasound for the early diagnosis of thyroid cancer: a systematic review and meta-analysis [Aplicación y limitaciones de la ecografía para el diagnóstico precoz del cáncer de tiroides: una revisión sistemática y metanálisis]. *American Journal of Translational Research*, 17(8), 6556-6572. <https://doi.org/10.62347/BNBG2830>

- Mandeville P. (2005, julio-septiembre). Tema 9: El coeficiente de correlación intraclase (ICC). *Ciencia UANL*, 8(3), 414-416. <https://www.redalyc.org/pdf/402/40280322.pdf>
- Moragas, A., Monfà, R., García-Sangenís, A., & Llor, C. (2026). Accuracy of leukocyte esterase and nitrite tests for diagnosing bacteriuria in older adults: A systematic review and meta-analysis [Precisión de las pruebas de esterasa leucocitaria y nitrito para el diagnóstico de bacteriuria en adultos mayores: una revisión sistemática y metanálisis]. *Clinical Microbiology and Infection*, 32(1), 19-29. <https://doi.org/10.1016/j.cmi.2025.08.027>
- Pimenta, S., Pissarra, S., Soares, P., Azevedo, I., & Pereira-Nunes, J. (2025). Serum biomarkers in the early detection of necrotizing enterocolitis: a systematic review [Biomarcadores séricos en la detección temprana de la enterocolitis necrotizante: una revisión sistemática]. *Journal of Perinatal Medicine*, 53(8). <https://doi.org/10.1515/jpm-2025-0180>
- Pita Fernández, S., & Pértegas Díaz, S. (2003). Pruebas diagnósticas: Sensibilidad y especificidad. *Cadernos de Atención Primaria*, (10), 120-124. <https://www.fisterra.com/formacion/metodologia-investigacion/pruebas-diagnosticas-sensibilidad-especificidad/>
- PRISMA. (2025). *PRISMA-Diagnostic Test Accuracy* [Precisión de la prueba diagnóstica PRISMA]. <https://www.prisma-statement.org/dta>
- Radmard, M., Gad, M., Sheikhy, A., Dagher, R., Yedavalli, V., & Luna, L. P. (2025). Functional Imaging in Alzheimer's Disease: A Systematic Review and Meta-Analysis of ASL-MRI and FDG-PET [Imágenes funcionales en la enfermedad de Alzheimer: una revisión sistemática y metanálisis de ASL-MRI y FDG-PET]. *Journal of Neuroimaging*, 35(5), Article e70080. <https://doi.org/10.1111/jon.70080>
- Rendón-Macías, M. E., Valenzuela, M., & Villasís-Keever, M. A. (2020). Sesgos en los estudios de pruebas de diagnóstico: implicación en la estimación de la sensibilidad y especificidad. *Revista Alergia México*, 67(2), 165-173. <https://doi.org/10.29262/ram.v67i2.771>
- Webster, K. E., Parkhouse, T., Dawson, S., Jones, H. E., Brown, E. L., Hay, A. D., Whiting, P., Cabral, C., Caldwell, D. M., & Higgins, J. P. (2024). Diagnostic accuracy of point-of-care tests for acute respiratory infection: a systematic review of reviews [Precisión diag-

nóstica de las pruebas en el punto de atención para la infección respiratoria aguda: una revisión sistemática de revisiones]. *Health Technology Assessment*. <https://doi.org/10.3310/JLCP4570>

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M., Sterne, J. A., Bossuyt, P. M., & QUADAS-2 Group. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies [QUADAS-2: una herramienta revisada para la evaluación de la calidad de los estudios de precisión diagnóstica]. *Annals of Internal Medicine*, 155(8), 529-536. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>

Capítulo 2. La magnitud y distribución de una enfermedad: síntesis de la literatura de estudios descriptivos

1. Características de los estudios descriptivos

Los estudios descriptivos son investigaciones observacionales centradas en conocer la ocurrencia de un evento en salud y su distribución poblacional, por ello cumplen un rol fundamental en la investigación médica; son la base de la pirámide en la jerarquía de la evidencia y con ellos inician todas las áreas nuevas de investigación alrededor de una enfermedad. Los investigadores que desarrollan estudios descriptivos dan cuenta de la frecuencia, los factores asociados⁵ con una enfermedad, describen las características de presentación de la enfermedad, las particularidades de las personas afectadas y generan hipótesis sobre los factores causales. En estos estudios, el investigador no interviene manipulando-controlando variables, simplemente observa ciertas características de la población a medida que ocurren “naturalmente” y registra los datos relevantes (Munn et al., 2017).

Los estudios descriptivos incluyen los reportes de caso, las series de casos⁶, los estudios de prevalencia, incidencia (en eventos agudos), endemias, brotes, encuestas de morbilidad y los estudios ecológicos. Entre ellos, se destacan los estudios de prevalencia, transversales o *cross-sectional*⁷, porque generan información sobre la magnitud de una enfermedad y la presencia o ausencia de una exposición en un momento específico; además, se comparan diferentes subgrupos, conformados en función de las variables independientes del estudio, con respecto a la prevalencia de la enfermedad. **En es-**

⁵ Los factores asociados corresponden a la variación en la magnitud (distribución) de la enfermedad según características de la población de estudio; son diferentes de los factores de riesgo que corresponden a variables que tienen una relación causal con la enfermedad y se estudian en investigaciones analíticas longitudinales (el factor de riesgo debe anteceder al evento).

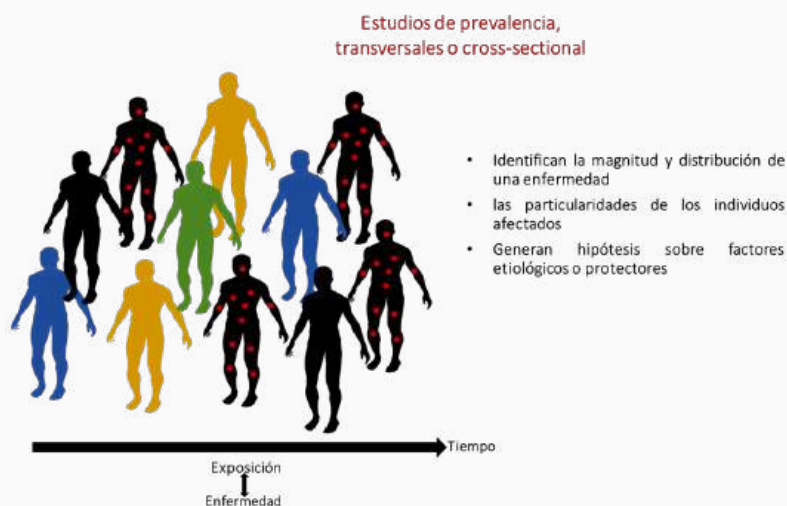
⁶ Se podrían diferenciar los estudios descriptivos de los estudios de casos y las series de casos, enfatizando que los dos últimos buscan caracterizar eventos de salud inusuales o experiencias clínicas únicas, generalmente en grupos pequeños de 1-10 individuos, en los estudios de casos, y más de 10 en las series de casos, privilegiando un enfoque individualizado que resalte las particularidades y variaciones de cada sujeto. El enfoque individual y la utilidad para documentar condiciones inusuales marcarían las principales diferencias con los estudios descriptivos.

⁷ Se debe tener presente que algunos autores traducen al español el término “cross-sectional” como estudios transversales analíticos, con dos o más grupos de estudio; es este caso no se incluirían en los estudios descriptivos. Para este libro, el término “cross-sectional” se utiliza como aparece en los términos MeSH y DeCS; es decir, como antónimo de los estudios longitudinales, donde lo central no es el número de grupos o poblaciones estudiadas, sino la temporalidad de la medición de la variable dependiente (enfermedad o evento en salud) y las independientes (exposición).

tos estudios, la enfermedad y la exposición se miden de manera simultánea, así que no es posible saber si una de ellas antecede a la otra; es decir, no es apropiado establecer relaciones causales. En este sentido, la pregunta sobre qué fue primero, el huevo o la gallina, no es posible responderla en un estudio de prevalencia, con excepción de aquellos casos en que la exposición no cambia en el tiempo y puede establecerse teóricamente que esta sí antecede a la enfermedad; por ejemplo, el sexo, algunas variables genéticas, algunos casos de exposiciones ambientales, entre otros (ver Figura 14).

Figura 14

Representación gráfica de un estudio de prevalencia, transversal o cross-sectional



Nota. Elaboración propia.

Entonces, la prevalencia es definida como el número de individuos con una enfermedad presente en la población en un tiempo específico, dividido el número total de individuos en esa población y en ese momento específico⁸. La prevalencia se puede clasificar como prevalencia de punto (número total de enfermos existentes en una población en un momento) y prevalencia de período (número total de enfermos existentes en una población durante un período de tiempo). En este punto es importante aclarar que lo central de un caso prevalente radica en el hecho de no estar interesado en diferen-

⁸ Esta definición se focaliza en mostrar el numerador y denominador del cálculo, mostrando que lo importante del concepto es remitir a una proporción de afectados por el evento. Si solo se toma la proporción, esta definición también aplicaría para los conceptos de frecuencia de la enfermedad o seroprevalencia; sin embargo, en algunos contextos se han establecido diferencias en estos términos, que se explican en páginas siguientes.

ciar si este es nuevo o antiguo para la enfermedad, infección o evento en salud estudiado; por ello, es más utilizado en eventos crónicos, a diferencia de los casos incidentes que sí deben demostrar que es nuevo, sea por tratarse de un evento agudo o por disponer de al menos dos mediciones en el tiempo (con una prueba previa negativa).

En esta modalidad de investigación se abordan preguntas como: ¿cuántas personas tienen la enfermedad?, ¿quiénes contraen la enfermedad? y ¿dónde está ocurriendo la enfermedad? Para dar respuesta a estas preguntas, es fundamental hacer un proceso riguroso en la selección de los sujetos de estudio. **Para que los resultados de un estudio de prevalencia puedan generalizarse a toda la población, es necesario incluir una muestra representativa de esa población a partir de un muestreo aleatorio.** Si el estudio no se realiza con una muestra, sino que se incluye a toda la población, entonces estos se convierten en un censo y, en los casos en los que la selección de los sujetos de estudios se haga a conveniencia de los investigadores, los resultados no pueden generalizarse más allá de la muestra estudiada. En este último escenario, resulta de gran importancia la realización de una descripción exhaustiva de la población de estudio (de las variables independientes de la población) con el fin que el lector pueda establecer a qué tipo de grupos podrían extrapolarse los resultados.

Los estudios de prevalencia son particularmente valiosos para monitorear las tendencias en la carga y la aparición de enfermedades, para orientar políticas sanitarias, focalizar recursos y orientar intervenciones. Además, como se mencionó antes, pueden ser útiles para formular hipótesis sobre aspectos etiológicos, factores de riesgo o factores protectores en el desarrollo y progresión de una enfermedad (Munn et al., 2017; Rothman et al., 2008).

En adición a las diferencias conceptuales y metodológicas de los estudios de prevalencia frente a los de incidencia, en algunos escenarios académicos también es frecuente diferenciar los primeros de los estudios de seroprevalencia y de frecuencia. Los estudios de seroprevalencia cumplen con las mismas características descritas previamente para los estudios de prevalencia, pero la medición del evento se hace mediante detección de anticuerpos y otras pruebas serológicas que permiten dar cuenta de infecciones subclínicas, latentes o estado de portador o expuesto; por ello, también se denominan estudios de seropositividad o sero-reactividad (la principal diferencia con el concepto de prevalencia⁹ sería el numerador; es

⁹ En este caso, la definición de prevalencia, además de aludir el cálculo de la proporción, implica tener en el denomi-

muy útil principalmente para eventos infecciosos en los cuales es importante diferenciar los expuestos (que se refiere solo al contacto con el agente) de los infectados (en quienes el agente penetra al cuerpo y se replica).

El término de estudios de frecuencia generalmente se sugiere para diseños similares a los de prevalencia, pero que cumplen con una de estas dos condiciones: no se realizan en toda la población o una muestra representativa de ella, y no se aplica la prueba estándar (también llamadas “de oro”) para la detección del evento; sin embargo, se debe considerar que los tesauros MeSH y DeCS no dan cuenta de estas características e incluyen la frecuencia como un sinónimo de prevalencia; además, en los libros tradicionales de epidemiología se usa como sinónimo de ocurrencia o magnitud; es decir, la frecuencia subsume la prevalencia y la incidencia. En la Tabla 7 se describen las principales características de este tipo de estudios.

Tabla 7
Características distintivas de un estudio transversal de prevalencia

Característica	Estudio Transversal de Prevalencia
Objetivo principal	Describir la prevalencia de una enfermedad y sus factores asociados en una población y en un momento específico de tiempo.
Tipo de estudio	Observacional, descriptivo.
Dirección del estudio	Transversal; se realiza en un solo punto en el tiempo, recopilando datos simultáneamente sobre la exposición y la enfermedad.
Selección de participantes	Se selecciona una muestra representativa de la población en el momento de la medición.
Muestra	La muestra debe ser representativa para hacer generalizaciones válidas sobre la población.
Medición de la exposición	La exposición y la enfermedad se miden simultáneamente. No es posible establecer la antecendencia de la exposición (no se establecen relaciones causales)
Costo	Relativamente bajo, debido a que los datos se recopilan en un solo punto de tiempo.
Tiempo requerido	Corto, ya que no requiere seguimiento a largo plazo.
Principal medida de desenlace	Prevalencia, que se refiere a la proporción de personas con la enfermedad en la población, en el momento de estudio.

nador la población expuesta o una muestra representativa de ella (sujetos elegidos de manera probabilística) y que el numerador sea calculado con pruebas estándar o de referencia para el evento.

Ventajas	Rápido y económico; permite estimar la carga de enfermedad en una población; adecuado para estudiar enfermedades comunes y factores de riesgo.
Desventajas	No permite establecer relaciones causales.

Nota. Elaboración propia.

2. Sobre la validez interna y externa en los estudios descriptivos

En adición a la caracterización anterior y previo a la presentación de los detalles operativos para sistematizar la evidencia generada con este tipo de estudios, es necesario conocer algunas fuentes de error en su ejecución que pueden afectar su validez externa (posibilidad de generalización de los resultados) e interna (medir sin errores o control de sesgos). Esto resulta necesario porque permite mejorar la formulación y ejecución de este tipo de estudios, analizar críticamente el tipo de evidencia que genera, y en el caso de las revisiones sistemáticas permite evaluar la calidad metodológica de los estudios incluidos y, subsecuentemente, garantizar la validez de las conclusiones derivadas de dicha sistematización.

Es importante resaltar que en los estudios transversales se están identificando casos prevalentes. Tales casos pueden no ser representativos de todas las personas con la enfermedad en la población, ya que las personas que mueren poco después del diagnóstico o se recuperan rápidamente tienen menos probabilidades de ser identificadas como enfermas. Así, su utilidad es limitada en enfermedades de evolución rápida (curación o muerte), ya que pueden derivar en una subestimación del evento (Aschengrau & Seage, 2020; Celentano et al., 2024).

Por otro lado, la presencia o ausencia de exposición y enfermedad se determinan en un mismo momento en el tiempo en cada participante en el estudio, así que no es posible establecer una relación temporal entre la exposición y el inicio de la enfermedad. **Sin información sobre las relaciones temporales no es posible establecer, por ejemplo, si el sedentarismo antecede al síndrome metabólico o se da posterior a este, en cuyo caso se presentaría el fenómeno de “causalidad inversa”.** En últimas, dado que no hay certeza sobre si la exposición precede el desarrollo de la enfermedad, las asociaciones no pueden reflejar relaciones causales (Celen-

tano et al., 2024).

Pese a lo anterior, se debe considerar que las relaciones causales son muy importantes para orientar intervenciones, pero en el marco de los estudios de prevalencia, la correlación del evento y algunas exposiciones son útiles para conocer la distribución de la enfermedad, hacer focalización de esfuerzos investigativos posteriores, entre otros usos¹⁰.

Aunado a lo anterior, la prevalencia de una enfermedad podría estar influenciada por la definición del caso; por ejemplo, el síndrome metabólico está compuesto por obesidad abdominal, hipertrigliceridemia, colesterol HDL bajo, presión arterial elevada y alteración en la regulación de la glucosa. Para la IDF (*International Diabetes Federation*) el diagnóstico del síndrome se hace con la presencia de obesidad abdominal más otros dos factores alterados, mientras que para la *National Cholesterol Education Program Adult Treatment Panel III* (ATP III), el diagnóstico se hace con la presencia de tres de los cinco factores enunciados (Carvajal Carvajal, 2017). Por supuesto, la adherencia a uno u otro criterio puede condicionar la frecuencia hallada.

Del mismo modo, las mediciones pueden estar influenciadas por las técnicas utilizadas o por las personas que las realizan; por ejemplo, se pueden obtener diferentes proporciones de diabetes si el diagnóstico se hace con hemoglobina glicosilada, glicemia en ayunas, la prueba de tolerancia oral a la glucosa o una prueba casual de glucosa plasmática.

Otra limitación de los estudios transversales está relacionada con el sesgo de supervivencia. El sesgo de supervivencia consiste en que una exposición puede empeorar una condición clínica al punto que afecta la supervivencia de los individuos expuestos, y como resultado la exposición a ese factor se encontrará con menor frecuencia en los casos prevalentes, no por su ausencia, sino porque los individuos expuestos mueren con mayor facilidad. Este tipo de sesgo también se conoce como sesgo de prevalencia-incidencia (Celentano et al., 2024). En la Tabla 8 se resumen los potenciales sesgos de los estudios descriptivos.

A pesar de sus limitaciones, los estudios transversales son ampliamente utilizados, y a menudo son los primeros estudios realiza-

¹⁰ No es riguroso pensar que solo las relaciones causales importan o que las no causales resultan espurias, lo importante es evitar extremos: no incurrir en la falacia de correlación (interpretar todas las correlaciones son causales) ni pretender que solo las relaciones causales son útiles, dado que no siempre se requiere predecir o aislar el mecanismo por el cual una variable afecta a otra.

dos porque permiten generar hipótesis que luego puedan evaluarse en diseños más complejos en los que se puede establecer con mayor certeza la secuencia temporal entre la exposición y el evento (Celementano et al., 2024). Los estudios transversales tienen la ventaja de que generan una gran cantidad de información con una baja inversión de tiempo y dinero. Además, cuando se basan en una muestra aleatoria de la población general, arrojan resultados altamente generalizables.

Tabla 8
Listado de potenciales sesgos en estudios descriptivos

Etapa del estudio		Control de potenciales sesgos
Diseño del estudio	Definición vigente del evento	Búsqueda adecuada de la evidencia disponible
	Búsqueda adecuada de la evidencia disponible	
Selección	Selección aleatoria	Aplicación rigurosa de criterios de inclusión y exclusión
	Aplicación rigurosa de criterios de inclusión y exclusión	
Recolección	Aplicar encuestas válidas	Aplicar encuestas válidas Elegir la prueba o criterio clínico válido para medir el desenlace Estandarización de mediciones (de los sujetos que participan de ellas y de los equipos empleados) Realizar las mediciones biológicas en laboratorios con control de calidad interno y externo Evitar el sesgo de cortesía y el de fatiga con buenas selección y entrenamiento del personal que hará el trabajo de campo Si es posible consultar datos en la historia clínica (o una fuente secundaria confiable) para evitar el sesgo de memoria
	Elegir la prueba o criterio clínico válido para medir el desenlace	
	Estandarización de mediciones (de los sujetos que participan de ellas y de los equipos empleados)	
	Realizar las mediciones biológicas en laboratorios con control de calidad interno y externo	
	Evitar el sesgo de cortesía y el de fatiga con buenas selección y entrenamiento del personal que hará el trabajo de campo	
Análisis	Si es posible consultar datos en la historia clínica (o una fuente secundaria confiable) para evitar el sesgo de memoria	Elegir los estadísticos o mediciones epidemiológicas correctas para cada objetivo específico
	Elegir los estadísticos o mediciones epidemiológicas correctas para cada objetivo específico	

Nota. Elaboración propia.

En este punto, es importante aclarar que en algunos textos de epidemiología la confusión es clasificada como un sesgo; no obstante, el fenómeno de la confusión debe separarse de los sesgos, porque a diferencia de las relaciones espurias resultantes de un sesgo, las relaciones que se establecen con las variables confusoras son reales. En este sentido, se recomienda que la evaluación de la confusión se haga con un análisis multivariado o un análisis estratificado en lugar de plantearse como un sesgo a controlar (Szklo & Nieto, 2014).

3. Indicaciones para sistematizar la evidencia de los estudios descriptivos y elegir cuándo combinar los resultados

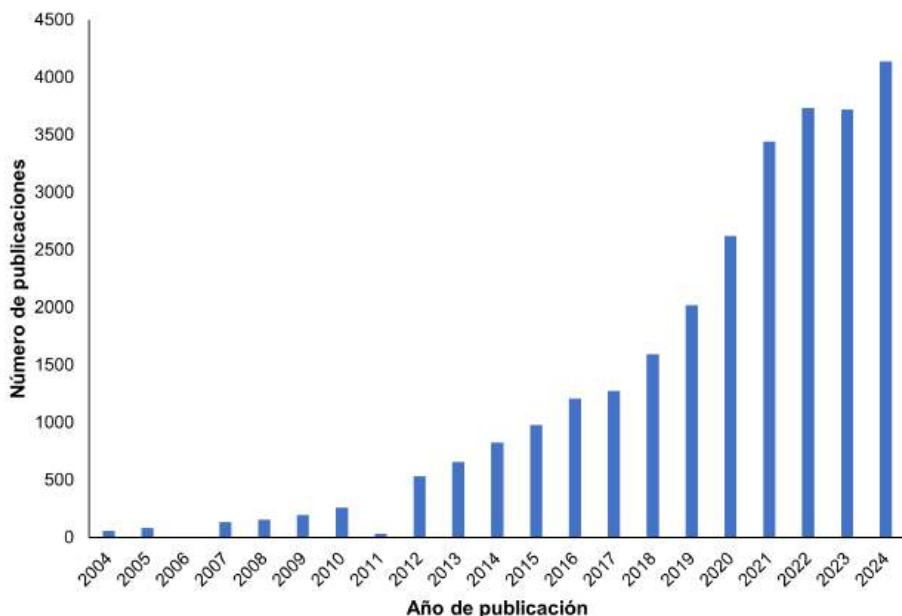
La síntesis de la literatura relacionada con estudios transversales consiste en determinar la frecuencia y distribución de una enfermedad en un grupo poblacional, un área geográfica particular y en un periodo de tiempo específico. El objetivo de estas revisiones es describir el problema de salud (qué), los afectados por él (quién), así como la ubicación (dónde) y el período de tiempo (cuándo) en el que ocurre. Teniendo en cuenta lo anterior, los datos que deben sintetizarse en una revisión sistemática de la literatura sobre estudios de prevalencia o transversales son básicamente proporciones; es decir, porcentajes de una población que experimenta una enfermedad.

Aunado a ello, también es posible desarrollar estas investigaciones con desenlaces continuos, en estudios en los que se aplican escalas y que arrojan puntuaciones sobre el estado de una población. Ese es el caso de las investigaciones sobre calidad de vida relacionada con la salud, mediciones metabólicas (valores de analitos relacionados con el perfil lipídico, renal, hepático, etc), hematológicas (parámetros del hemoleucograma, ferritina, etc), inmunológicas, entre otras.

En este sentido, diferentes investigadores han emprendido revisiones sistemáticas de estudios transversales. Así, una búsqueda en PubMed, con los términos (systematic review[Title]) AND (prevalence[Title/Abstract]) pone de manifiesto cerca de 10.000 publicaciones, solo en los últimos 20 años (ver Figura 15).

Figura 15

Frecuencia de publicaciones en PubMed sobre revisiones sistemáticas de estudios de prevalencia entre los años 2004 y 2024



Nota. Elaboración propia.

Entre esas publicaciones, se encuentran síntesis de la literatura de estudios de prevalencia de síndrome metabólico en pacientes con trastorno bipolar (Salari et al., 2025), pobre calidad del sueño en pacientes con dispepsia (Netyam et al., 2025), prevalencia del virus del papiloma humano y genotipos entre mujeres con VIH en África (Amare et al., 2025), tuberculosis en población privada de la libertad (Moreira et al., 2019), depresión en estudiantes de medicina (Puthran et al., 2016), entre muchas otras enfermedades. Además, este método se ha aplicado a tópicos relacionados con los servicios de salud que no necesariamente se asocian con una enfermedad; por ejemplo, se han abordado problemas como la prevalencia de bullying entre estudiantes de medicina (Li et al., 2025), de acoso y violencia sexual entre mujeres (Debnath et al., 2025), prevalencia de conducción bajo los efectos del alcohol (Pelletti et al., 2024), infecciones asintomáticas por el virus del zika (Haby et al., 2018) y prevalencia de errores médicos (Vaziri et al., 2019), entre otros.

Entre las razones para realizar una síntesis de la literatura de estudios transversales se encuentra que generan información sobre la carga global de una enfermedad, la descripción de su distribución geográfica, la variación entre subgrupos y países; permite monitorear cambios en el tiempo que revelen enfermedades emergentes, reemergentes o en disminución; son de gran utilidad cuando no es práctico realizar estudios multicéntricos o de escala internacional;

y, además, son de particular importancia cuando los estudios individuales están limitados por tamaños de muestra pequeños, con baja potencia estadística y escasa validez externa (Munn et al., 2017).

En este orden de ideas, al plantear el problema de investigación que sustenta la necesidad de realizar una revisión sistemática de estudios descriptivos se deben tener presente las siguientes consideraciones:

1. Delimitar la población: si es algún grupo etario (niños, adolescentes, adultos; o si no es necesario restringir por edad sustentarlos), sujetos de una ocupación dada (militares, trabajadores sexuales, profesionales de la salud) o en una condición especial (presos, usuarios de drogas).
2. Explicitar los elementos que inciden en la variación del evento a estudiar (factores asociados o de riesgo).
3. Justificar la elección de la escala espacial: si se trata de la evidencia de un único país, una región, un continente o incluso si se abordará una problemática en el ámbito mundial.
4. Argumentar la delimitación temporal (un año, un lustro, una década, etc.) o si no es necesario hacer restricciones de tiempo.

Estos elementos deben explicitarse en la pregunta de investigación que orienta el diseño del protocolo de búsqueda y selección de los estudios. Si bien no se trata de la clásica estructura de una pregunta PICO (Población, Intervención, Comparación, Desenlace) propia de estudios clínicos (y que muchos revisores la piden para todo tipo de revisiones sistemáticas), los elementos anteriores dan cuenta de una estructura similar que podría denominarse pregunta PEET (Población, Exposiciones, Espacio, Tiempo). Para cada uno de los elementos de esta pregunta, el problema debería mostrar heterogeneidad de la evidencia disponible, baja precisión de las estimaciones (por bajos tamaños de muestra o alta heterogeneidad del fenómeno de estudio) o evidencia no concluyente.

3.1 Características del protocolo de búsqueda y selección de estudios en la síntesis de estudios transversales

3.1.1 La estrategia de búsqueda

El protocolo de búsqueda es clave para garantizar la reproducibilidad y exhaustividad de la síntesis de la literatura. En el protocolo de un estudio que pretenda sistematizar estudios de prevalencia es

necesario explicitar la enfermedad o el problema de salud, la medida epidemiológica utilizada para cuantificar su frecuencia (Ej., prevalencia), la población de interés, la delimitación geográfica (en los casos que aplique) y una delimitación temporal en la que se especifique el periodo de estudio (Munn et al., 2017).

Para la enfermedad o problema de salud de interés se deben utilizar sinónimos, preferiblemente identificados en plataformas de lenguaje controlado como los términos Decs (descriptores de ciencias de la salud) y/o Mesh (*Medical Subject Headings*), y en combinación con el operador booleano OR. Así, por ejemplo, si el tópico de interés consiste en determinar la prevalencia de síndrome metabólico, la estrategia de búsqueda debería contener los términos: Síndrome metabólico OR síndrome metabólico cardiovascular OR Síndrome X Dismetabólico OR Síndrome X Metabólico Síndrome X de Reaven OR Síndrome X de Resistencia a la Insulina.

En cuanto a la medida epidemiológica utilizada para cuantificar la frecuencia de la enfermedad, el procedimiento es similar. En este sentido, al término prevalencia deben sumarse los términos: brotes OR endemia OR epidemia OR frecuencia OR morbilidad OR ocurrencia OR vigilancia. Sin embargo, diversos temas de salud no tienen indización en los tesauros, en cuyo caso se recomienda seleccionar los términos mediante una “bola de nieve” que consiste en identificar artículos de revisión bibliográfica del tema de interés, hallar los términos clave y sus sinónimos, y con estos realizar una nueva búsqueda de revisiones en todas las bases de datos de la revisión, con el fin de hallar nuevos sinónimos; el proceso termina cuando se hallen nuevas palabras (Schlosser et al., 2006).

En tercer lugar, se debe especificar la población de interés, en cuyo caso podrían ser adolescentes, estudiantes universitarios, población general, adultos mayores, entre otros, cada uno con los equivalentes en los bases de datos de lenguaje controlado. En este punto es importante precisar que, dependiendo de la naturaleza de la revisión, de las opciones de filtro de las bases de datos consultadas y de la proliferación de investigaciones en el tema de interés, la delimitación poblacional, espacial y temporal puede hacerse desde la estrategia de búsqueda o en la etapa de aplicación de los criterios de inclusión. Por ejemplo, en un estudio con el objetivo de sistematizar las investigaciones sobre tuberculosis resistente a los antibióticos en países de las Américas, la delimitación espacial debe hacerse al momento de aplicar los criterios de inclusión, de lo contrario, la estrategia de búsqueda debería dar cuenta de los más de 30 países de las

Américas, lo que afecta la sensibilidad de la búsqueda. Por su parte, un estudio que se plantee el objetivo de sistematizar la prevalencia de obesidad en niños y adolescentes, los “niños” y “adolescentes” deberían incluirse en la estrategia de búsqueda para aumentar la especificidad de la búsqueda.

Una vez aclarado el punto anterior, es necesario especificar las fuentes de información. En todas las síntesis de la literatura el número de fuentes consultadas es fundamental para garantizar la exhaustividad, así que deben incluirse tantas bases de datos como sea posible. Para los propósitos del libro, los ejemplos se realizarán con tres bases de datos: Scielo (reúne gran parte de las publicaciones científicas en revistas latinoamericanas), *PubMed* (permite consultar gran parte de la literatura científica en el ámbito internacional) y el buscador abierto *Google Scholar*; sin embargo, se recomienda que las revisiones sistemáticas agreguen a estas fuentes las siguientes bases de datos: EMBASE, CINAHL (EMCare), Cochrane/Campbell Collaboration, Scopus, ScienceDirect, Web of Science, dado que su cobertura temporal y espacial garantiza exhaustividad en todo el mundo.

Para finalizar la estrategia de búsqueda es necesario especificar si las palabras seleccionadas serán restringidas a alguna parte del texto, por ejemplo, al título, al resumen, a las palabras clave o si no habrá restricción. Para ello, es importante tener en cuenta que una búsqueda restringida al título asegura una alta especificidad, de modo que seguramente los artículos que se recuperen tratarán el tópico de interés; sin embargo, la sensibilidad para detectar otros manuscritos que no tengan las palabras en el título, pero que se refieran al tema de estudio, es baja. Por otro lado, al no aplicar restricciones, se asegura una alta sensibilidad, pero se recuperarán grandes cantidades de textos que aborden el tópico de estudio de manera marginal, así que la especificidad será muy baja (lo que también redundará en problemas para garantizar la reproducibilidad). En este sentido, una manera para conciliar el balance entre sensibilidad y especificidad, o entre exhaustividad y reproducibilidad, podría lograrse restringiendo las estrategias de búsqueda al resumen y, en los casos en los que los motores de búsqueda lo permitan, al título/resumen/palabra clave.

Una vez establecidas las palabras clave, la combinación con los operadores booleanos y la selección de las bases de datos, se debe hacer la búsqueda y registrar los resultados del modo que se presenta en la Tabla 9.

Tabla 9
Frecuencia absoluta de los artículos identificados en cada base de datos y con cada estrategia de búsqueda

PubMed Title/Abstract				
AND	Síndrome metabólico	síndrome metabólico cardiovascular	Síndrome X	Resistencia a la Insulina
Prevalencia	n	n	n	n
Frecuencia	n	n	n	n
Ocurrencia	n	n	n	n
Endemia	n	n	n	n
Epidemia	n	n	n	n
Morbilidad	n	n	n	n
Vigilancia	n	n	n	n
Scielo Title/Abstract				
AND	Síndrome metabólico	síndrome metabólico cardiovascular	Síndrome X	Resistencia a la Insulina
Prevalencia	n	n	n	n
Frecuencia	n	n	n	n
Ocurrencia	n	n	n	n
Endemia	n	n	n	n
Epidemia	n	n	n	n
Morbilidad	n	n	n	n
Vigilancia	n	n	n	n
Google Scholar Title				
AND	Síndrome metabólico	síndrome metabólico cardiovascular	Síndrome X	Resistencia a la Insulina
Prevalencia	n	n	n	n
Frecuencia	n	n	n	n
Ocurrencia	n	n	n	n
Endemia	n	n	n	n
Epidemia	n	n	n	n
Morbilidad	n	n	n	n
Vigilancia	n	n	n	n

Nota. Elaboración propia.
 *Los términos Síndrome X Dismetabólico, Síndrome X Metabólico y Síndrome X de Reaven no se incluyen en la búsqueda porque están contenidos en el término síndrome X.

En algunos contextos se puede sugerir el diseño de una única estrategia de búsqueda utilizando la combinación de todos los tér-

minos con operadores booleanos como AND, OR, NOT, por ejemplo: ((“Prevalence”[Title/Abstract] OR “Frequency”[Title/Abstract] OR “occurrence”[Title/Abstract] OR “morbidity”[Title/Abstract]) AND “metabolic syndrome”[Title/Abstract]) OR “x syndrome”[Title/Abstract] OR “insulin resistance”[Title/Abstract]; sin embargo, entre más extensa sea la estrategia, mayor será la probabilidad de errores por inclusión u omisión de paréntesis o de operadores booleanos en lugares incorrectos, lo que afecta la reproducibilidad. Además, algunas bases de datos como Science-Direct no generan resultados reproducibles cuando la estrategia de búsqueda tiene más de ocho palabras, o difieren en sus sintaxis; por ejemplo, en PubMed no es necesario hacer las búsquedas con cada término MeSH, porque sus búsquedas las incluyen por defecto; pero esto no funciona igual en otras fuentes. Para lo ilustrar el último punto ver el siguiente ejemplo:

El investigador digita esta búsqueda en PubMed: *Search: inborn errors of immunity*. Pero esta fuente, internamente, busca lo siguiente (que puede ser diferente a otras fuentes):

(“inborn”[All Fields] OR “inborns”[All Fields]) AND (“error”[All Fields] OR “error s”[All Fields] OR “errorful”[All Fields] OR “errors”[All Fields]) AND (“immune”[All Fields] OR “immuned”[All Fields] OR “immunes”[All Fields] OR “immunisation”[All Fields] OR “vaccination”[MeSH Terms] OR “vaccination”[All Fields] OR “immunization”[All Fields] OR “immunization”[MeSH Terms] OR “immunisations”[All Fields] OR “immunizations”[All Fields] OR “immunise”[All Fields] OR “immunised”[All Fields] OR “immuniser”[All Fields] OR “immunisers”[All Fields] OR “immunising”[All Fields] OR “immunities”[All Fields] OR “immunity”[MeSH Terms] OR “immunity”[All Fields] OR “immunization s”[All Fields] OR “immunize”[All Fields] OR “immunized”[All Fields] OR “immunizer”[All Fields] OR “immunizers”[All Fields] OR “immunizes”[All Fields] OR “immunizing”[All Fields])

Translations

inborn: “inborn”[All Fields] OR “inborns”[All Fields]

errors: “error”[All Fields] OR “error’s”[All Fields] OR “errorful”[All Fields] OR “errors”[All Fields]

immunity: “immune”[All Fields] OR “immuned”[All Fields] OR “immunes”[All Fields] OR “immunisation”[All Fields] OR “vaccination”[MeSH Terms] OR “vaccination”[All Fields] OR “immunization”[All Fields] OR “immunization”[MeSH Terms] OR

“immunisations”[All Fields] OR “immunizations”[All Fields] OR “immunise”[All Fields] OR “immunised”[All Fields] OR “immuniser”[All Fields] OR “immunisers”[All Fields] OR “immunising”[All Fields] OR “immunities”[All Fields] OR “immunity”[MeSH Terms] OR “immunity”[All Fields] OR “immunization’s”[All Fields] OR “immunize”[All Fields] OR “immunized”[All Fields] OR “immunizer”[All Fields] OR “immunizers”[All Fields] OR “immunizes”[All Fields] OR “immunizing”[All Fields]

Warnings

inborn errors of immunity

Stop word: of

En este sentido, **se recomienda utilizar búsquedas simples y cortas que contengan, al menos: i) enfermedad o evento de interés y ii) palabra que delimite el tipo de estudio. Alternativamente, se puede agregar la población de interés (adultos, jóvenes), la prueba diagnóstica,** entre otros términos que ayuden a planear una búsqueda específica. Además de la estrategia de búsqueda, en el protocolo se deben enunciar los criterios que utilizarán los autores para incluir, excluir estudios y especificar cómo se extraerán y sintetizarán los datos.

3.1.2 Los criterios de inclusión

La definición de los criterios de inclusión dependerá, por supuesto, del tema de estudio; sin embargo, es posible enunciar algunas características generales que se deberían tener en cuenta en esta sección, como la verificación del tipo de estudio (transversal, cross-sectional, prevalencia), el grupo etario (niños, adolescentes, adultos, adultos mayores), población de interés (militares, población general, hospitalizados, ambulatorios) y la delimitación geográfica (país, región, continente). Es particularmente importante describir en detalle la condición de salud, enfermedad, síntoma, evento o factor de estudio, excluyendo el mayor número de interpretaciones posibles, de manera que, si interesa estudiar infecciones por *Chlamydia*, es necesario detallar si la infección de interés es la ocasionada por *Chlamydia trachomatis* o *Chlamydia pneumoniae*. También, se puede proporcionar información sobre cómo se medirá, diagnosticará o confirmará la afección. Finalmente, pueden incluirse algunos aspectos operativos como el idioma de los artículos y el año de publicación.

3.1.3 Los criterios de exclusión

La exclusión de los artículos es fundamental para restringir la presencia de sesgos, así que se deben tener en consideración aspectos como que no reporten el numerador y/o el denominador con el cual se hizo el cálculo de la prevalencia, tamaños de muestra muy bajos que sean más propios de una serie de casos, artículos con problemas de validez interna por problemas de sesgos¹¹, resultados incongruentes con el objetivo del estudio, aspectos relacionados con un pobre desempeño diagnóstico (baja sensibilidad, especificidad, valores predictivos) de la prueba utilizada para hacer la identificación del evento. Adicional a estos criterios, es posible que algunos artículos que pasaron la etapa de tamización con la lectura del resumen no estén disponibles en texto completo en las bases de datos a las que usted tiene acceso; por tanto, no será posible aplicar otros criterios de exclusión ni extraer la información. En estos casos, es necesario agregar la no disponibilidad del texto completo como un criterio de exclusión.

3.1.4 La extracción de la información

La extracción de la información de los artículos que pasaron las etapas anteriores deberá adaptarse a la estratificación de las variables de interés en los estudios incluidos. Tales variables deben dar cuenta de tópicos clásicos en la epidemiología como el tiempo, el lugar y la persona. En este sentido, el formulario debe contemplar el año de realización del estudio, la ciudad o país en la que se desarrolló la investigación, el número y las características de las personas incluidas como factores demográficos, geográficos, sociales y de otro tipo. También, se debe contemplar los criterios diagnósticos del evento, la prueba diagnóstica utilizada, el número de personas con el evento o la condición de interés, el tipo de muestreo y otras variables que permitan hacer cálculos de la prevalencia global y la prevalencia específica por subgrupos, que permitan identificar y evaluar la fuente de heterogeneidad en la magnitud del evento, y realizar meta-regresiones (ver sección de análisis de la información) en los

¹¹ Este aspecto genera discusión dado que, en la visión más tradicional (por ejemplo, la de Cochrane aplicada a revisiones de intervenciones médicas) es necesario excluir estudios por sus problemas de calidad metodológica y presencia de sesgos. Sin embargo, nuestra sugerencia es NO excluir por sesgos, sino tomar estos ítems de calidad metodológica como una variable más a extraer y analizar en los estudios incluidos. Tener los sesgos como variables, en lugar de criterio de elegibilidad, tiene la ventaja de permitir la identificación de áreas por mejorar en los estudios posteriores, y de evaluar si la calidad metodológica de cada estudio afecta el desenlace (no se puede asumir que en todos los casos que no se hizo un control de todos los posibles sesgos de un estudio observacional este factor altere el resultado global).

casos que sea posible.

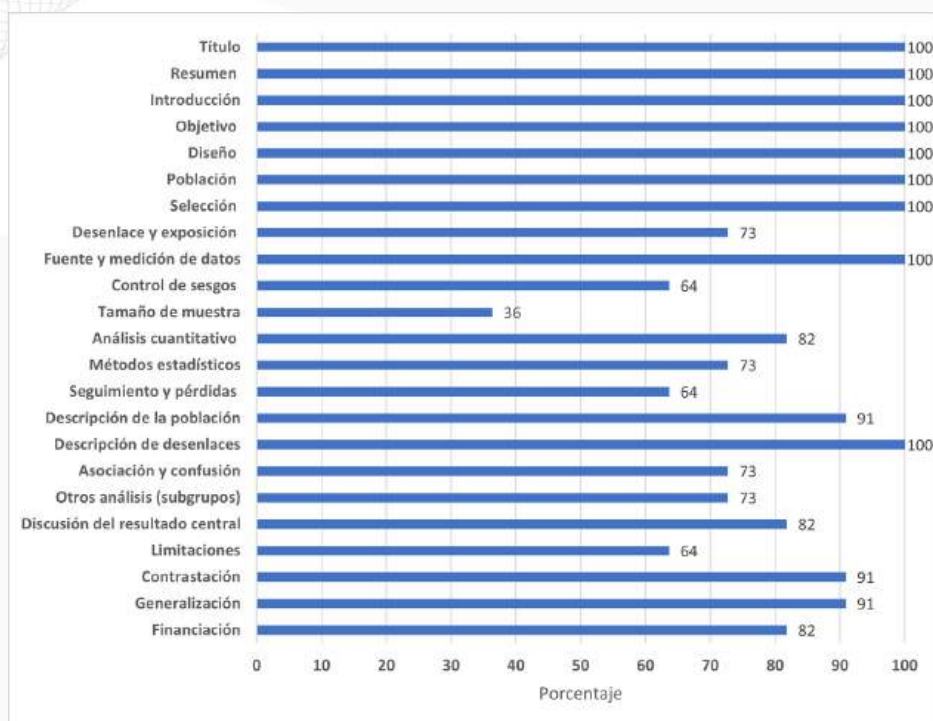
3.2 Evaluación de la calidad metodológica y riesgos de sesgos

En la literatura hay diversas herramientas para evaluar la calidad metodológica de estudios transversales (analíticos o descriptivos), entre ellos se encuentran la lista de verificación de evaluación crítica del instituto Joanna Briggs (JBI, 2020), la herramienta de evaluación para estudios transversales AXIS (Downes et al., 2016), y la guía STROBE (*STrengthening the Reporting of OBservational studies in Epidemiology*) para estudios transversales (STROBE, 2025). Aunque esta última es una guía para el reporte de esta modalidad de investigación y contempla asuntos editoriales relacionados con el título, el resumen o la financiación, también incluye asuntos clave de la calidad metodológica que dan cuenta de la validez interna y externa de las publicaciones, como la claridad en la definición de las variables, el cálculo del tamaño de la muestra y el muestreo, el control de sesgos, el análisis de la información, la presentación de resultados y la discusión. Entre estas tres herramientas, la lista de verificación del JBI es la más usada (Ma et al., 2020).

La evaluación de la calidad metodológica de los estudios debe ser realizada por dos revisores de manera independiente y luego reunirse para discutir los resultados de su evaluación crítica. Si existen discrepancias en una de las evaluaciones, estas pueden resolverse mediante consenso o se puede recurrir a un tercer revisor. En la Figura 16 se ejemplifican los resultados de la evaluación de la calidad metodológica de estudios transversales con la guía STROBE, y en la Tabla 10 se muestra el modo de presentación de los resultados de la evaluación de la calidad metodológica con los criterios de la guía del instituto Joanna Briggs. La representación en ambos casos puede ser en tabla o figura, siguiendo la proporción de estudios que cumple cada uno de los ítems contemplados (el denominador sería el número de estudios) o haciendo el análisis individual del cumplimiento de criterios de calidad de cada estudio (el denominador sería el número de criterios evaluados).

Figura 16

Evaluación de la calidad metodológica de los estudios según la guía STROBE para estudios transversales



Nota. Elaboración propia.

Se presenta la proporción de estudios que cumplen cada uno de los ítems contemplados en la guía, pero también se puede presentar como el porcentaje de criterios STROBE que cumple cada artículo.

Pese a lo anterior, se recomienda que en revisiones con muchos estudios (7 o más) se haga la presentación de la proporción de estudios que cumple cada criterio (esto permite que los interesados en el campo de estudio identifiquen puntos clave para mejorar sus diseños o reportes de investigación) y la tabla completa se envíe como material suplementario. Si bien la unidad de análisis de una revisión sistemática son los estudios y el resultado de calidad metodológica de cada investigación podría ser más importante para los lectores, en los análisis subsiguientes o en las meta-regresiones los interesados podrían ver la importancia de esa fase, al indicarse si el nivel de calidad metodológica de cada estudio es una fuente de heterogeneidad del efecto estimado o afecta el valor de la proporción de prevalencia (por ello sugerimos presentar la figura de la calidad metodológica en función de los criterios evaluados).

Tabla 10
Evaluación de la calidad metodológica de cada estudio según los criterios de la guía del instituto Joanna Briggs

Criterios	Autor 1	Autor 2	Autor 3	Autor 4
Se definieron los criterios de inclusión claramente	Si	No	No claro	Si
Se describieron en detalle los sujetos del estudio y el contexto	No	No	Si	Si
Se midió la exposición de forma válida y confiable	Si	No	Si	No
Se utilizaron criterios objetivos y estandarizados para medir el evento de interés	Si	No claro	Si	No
Se identificaron factores de confusión	Si	No claro	Si	No
Se establecieron estrategias para abordar los factores de confusión	Si	Si	Si	No claro
Se midieron los resultados de forma válida y confiable	Si	Si	Si	Si
¿Se utilizó un análisis estadístico apropiado?	Si	Si	Si	No

Nota. Elaboración propia.

3.3 Análisis de la información

El análisis de la información contempla dos secciones: una síntesis cualitativa y una síntesis cuantitativa. En la síntesis cualitativa se presenta un resumen narrativo con énfasis en informar las características de los estudios incluidos; para ello se pueden utilizar tablas, gráficos y otros diagramas que ayuden al lector a comprender la presentación de los resultados informados en cada estudio individual, con base en las variables extraídas (según los objetivos específicos de la revisión sistemática). También, se debe realizar el cálculo de frecuencias absolutas y relativas con sus intervalos de confianza del 95 %, ello permite dar cuenta de la distribución temporal de las publicaciones, proporción de individuos incluidos, proporción de individuos con el evento y distribución geográfica.

En cuanto a la distribución geográfica, en aquellas revisiones sistemáticas de alcance global, resulta particularmente ilustrativo la presentación de un mapa que refleje los resultados por país, región y

continente. Es importante explicitar que la síntesis cualitativa no tiene la connotación de técnicas de análisis cualitativo; en el contexto de las revisiones sistemáticas esta expresión se aplica para todo tipo de análisis (narrativo o estadístico) que no sea metanálisis (síntesis cuantitativa); es decir, se pueden realizar síntesis cualitativas usando estadísticas, siempre que sean diferentes a las técnicas meta-analíticas implicadas en la estimación de una medida combinada de los estudios individuales (análisis de heterogeneidad, sensibilidad y sesgo de publicación).

En su mayoría, las tablas comparten una estructura similar a la que se presenta a continuación (ver Tabla 11). En ella se describen los resultados de un estudio hipotético sobre prevalencia del virus del papiloma humano en hombres que tienen sexo con hombres. Observe que se presenta la población global, la prevalencia de infección con sus intervalos de confianza del 95 % y la prevalencia específica de genotipos oncogénicos y no oncogénicos del virus.

Tabla 11
Descripción de la prevalencia global de la infección por VPH y la prevalencia específica según el genotipo

Autor	Individuos incluidos	Individuos positivos	Prevalencia (IC95%)	Prevalencia específica	
				Oncogénicos	No oncogénicos
Autor 1	N	n	% (IC 95)	% (IC 95)	% (IC 95)
Autor 2	N	n	% (IC 95)	% (IC 95)	% (IC 95)
Autor 3	N	n	% (IC 95)	% (IC 95)	% (IC 95)
Autor 4	N	n	% (IC 95)	% (IC 95)	% (IC 95)
Autor 5	N	n	% (IC 95)	% (IC 95)	% (IC 95)
Autor 6	N	n	% (IC 95)	% (IC 95)	% (IC 95)
Autor 7	N	n	% (IC 95)	% (IC 95)	% (IC 95)
Autor 8	N	n	% (IC 95)	% (IC 95)	% (IC 95)

Nota. Elaboración propia.

Con respecto a la síntesis cuantitativa, se refiere a metanálisis en el que se evalúan los siguientes aspectos (Cardona-Arias et al., 2016):

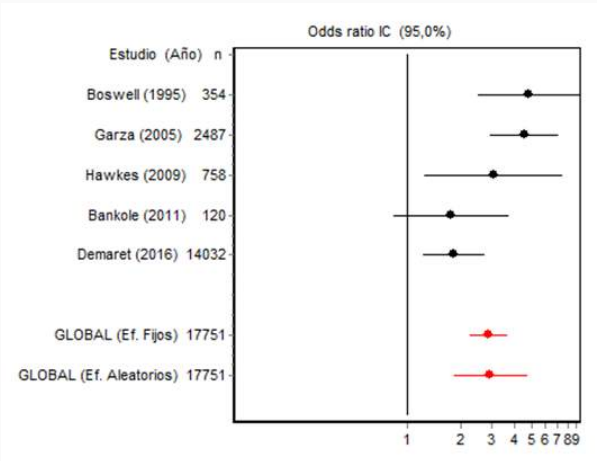
1. Heterogeneidad: determinación de la variabilidad o el error estándar de la proporción de prevalencia $[\sqrt{p(1-p)/n}]$, donde

- p es el porcentaje y n es el tamaño de la muestra]) entre los estudios incluidos. Esta se puede realizar con el I² (se define heterogeneidad con valores mayores a 0.5), con métodos gráficos (por ejemplo Galbraith) o pruebas de hipótesis (por ejemplo, Estadístico Q, DerSimonian Laird, etc.) donde la hipótesis nula plantea homogeneidad (con valores $p < 0,05$ se concluye que hay heterogeneidad). Si se presenta homogeneidad entre los estudios se realiza metanálisis de efectos fijos (el intervalo de confianza de la prevalencia combinada solo incluye la variación intra-estudios) y en casos de heterogeneidad metanálisis de efectos aleatorios (el intervalo de confianza de la prevalencia combinada incluye la variación intra e inter-estudios).
2. Sesgo de publicación: determina si los estudios incluidos tienden a confluir en ciertos valores de la prevalencia investigada, mostrando una suerte de preferencia de los editores científicos por un patrón (sesgado) de resultados. Generalmente se evalúa con el Funnel Plot esperando que los estudios presenten un tamaño del efecto (en este caso prevalencias) en amplio margen (con valores superiores e inferiores al efecto estimado) y variación de sus niveles de presión o error estándar. Este gráfico se puede complementar con pruebas de hipótesis, por ejemplo, la de Begg, cuya hipótesis nula indica que no se presenta este sesgo (valores $p < 0.05$ permiten afirmar que sí se presenta este sesgo).
 3. Análisis de sensibilidad para verificar la influencia de cada estudio en el resultado global, median el gráfico de influencia o tablas que presenten la medida prevalencia combinada de todos los estudios incluidos y la obtenida al eliminar cada estudio en fases sucesivas. Se concluye que existe buena sensibilidad en la medida combinada cuando todos los intervalos de traslapan (comparten valores).
 4. Tamaño del efecto o forest plot para exponer el valor puntual de la prevalencia y su intervalo de confianza del 95 % en cada estudio individual y la combinación por efectos fijos o aleatorios.

En los casos en que los estudios incluidos, además de reportar la prevalencia general, informen prevalencias específicas en algunos subgrupos, se podrían aplicar las siguientes opciones meta-analíticas: i) si todos los estudios incluidos reportan prevalencias específicas según la misma variable independiente (por ejemplo, prevalencia general de VPH y prevalencia específica por genotipo oncogénico y no oncogénico) se podría hacer un metanálisis para una diferencia,

razón de prevalencias o incluso odds ratio¹² (ver Figura 17), o aplicar una meta-regresión (metanálisis por subgrupos) por dicha variable independiente; ii) si el reporte de las prevalencias específicas por alguna variable independiente no se realizó de manera similar en todos los estudios incluidos, se podría hacer metanálisis de medidas indirectas, calculando de manera separada la medida combinada para cada grupo de prevalencias específicas. En el ejemplo del VPH se haría un metanálisis para la prevalencia de genotipos oncogénicos y otro independiente con los estudios que reportaron los no oncogénicos del VPH.

Figura 17
Forest plot de un metanálisis para la razón de odds (OR)



Nota. Elaboración propia.

En los estudios que presentan el evento desagregado por subgrupos (hombre y mujer) o cuando el desenlace es continuo y el estudio es transversal analítico, por ejemplo, las investigaciones que comparan la calidad de vida relacionada con la salud en individuos con enfermedades diferentes, es posible realizar un metanálisis para diferencia de medias. Ese metanálisis deberá incluir los resultados de la evaluación de la heterogeneidad, el sesgo de publicación y el

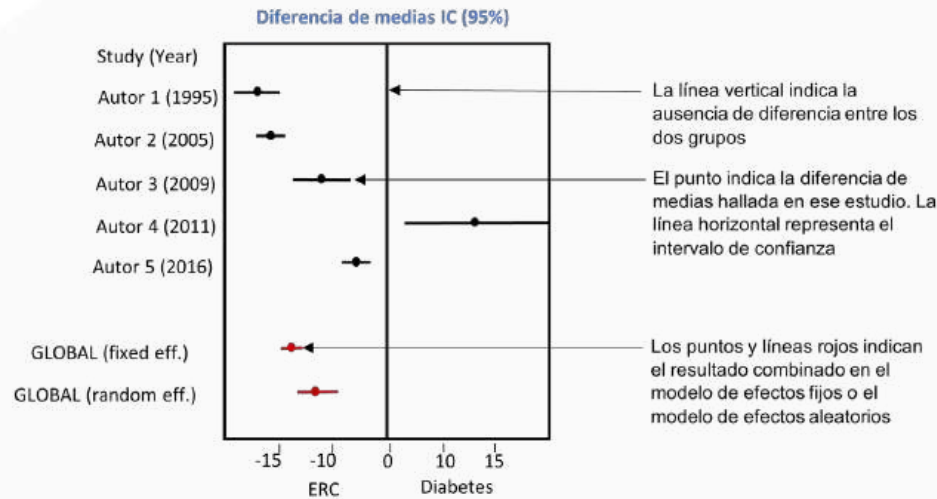
¹² Se podría hacer [metanálisis](#) para razones de prevalencias dividiendo la prevalencia de VPH oncogénicos / prevalencia de VPH NO oncogénicos, también se podría hacer [metanálisis](#) para la diferencia de proporciones (prevalencia de VPH oncogénicos - prevalencia de VPH NO oncogénicos). Incluso se podría hacer [metanálisis](#) para la razón de odds (OR), verificando que no se presente el problema de sobreestimación de la OR. Se considera que en enfermedades de baja prevalencia el valor de la OR y la razón de prevalencia es similar, pero en eventos de alta prevalencia la OR tiende a ser más elevada (sobreestima la fuerza de la asociación) que la razón de prevalencias. Algunos han indicado que estas dos razones cambian para prevalencias mayores al 20% y otros indican que incluso estas diferencias se dan para eventos con prevalencia o incidencia mayor al 10%.
https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0213-91112003000100012
<https://pmc.ncbi.nlm.nih.gov/articles/PMC3348192/>

forest plot como resultado central.

El forest plot para un estudio hipotético que tenga como objetivo comparar el impacto en la calidad de vida relacionada con la salud (CVRS) de pacientes con enfermedad renal crónica (ERC) y pacientes con diabetes deberá reflejar la diferencia de medias en la CVRS entre estos grupos en cada estudio individual y también la diferencia de medias combinando todas las publicaciones. Así, por ejemplo, se puede identificar que los pacientes con ERC tienen una CVRS inferior que la de los pacientes con diabetes en cerca de 13 puntos (en una escala de 0 a 100) (ver Figura 18).

Figura 18

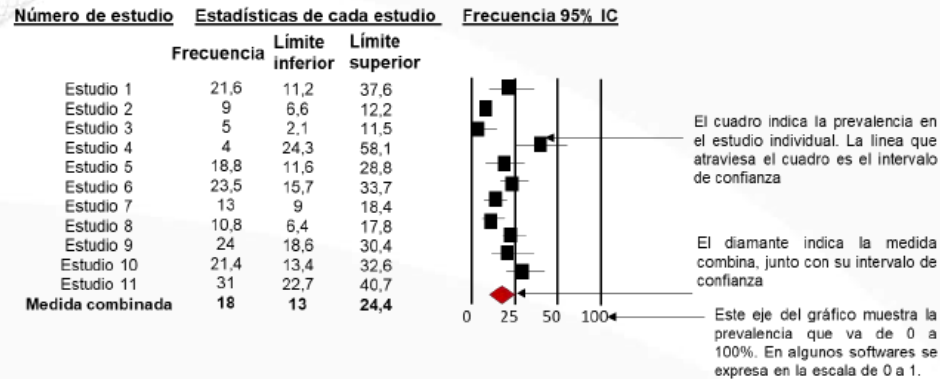
Forest plot para un estudio transversal de diferencia de medias



Nota. Elaboración propia.

Por su parte, en un estudio hipotético que tenga por objetivo sistematizar la prevalencia de síndrome metabólico en conductores de servicio público en el ámbito mundial, deberá reflejar la prevalencia hallada en cada estudio individual, junto con su intervalo de confianza; además, la prevalencia combinando todas las publicaciones. En este ejemplo, la prevalencia en los estudios individuales fluctuó entre el 4 % y el 31 %, mientras que la medida combinada fue de 18 % con un intervalo de confianza del 13 al 24,4 % (ver Figura 19).

Figura 19
Forest plot para estudio transversal de prevalencia



Nota. Elaboración propia.

4. Ejemplo práctico

Problema de investigación: el lupus eritematoso sistémico es una enfermedad autoinmune que tiene un importante componente inflamatorio. Estudios previos han sugerido con el colesterol HDL tiene un efecto antiinflamatorio y que podría estar disminuido en los pacientes con lupus; sin embargo, cuando se comparan los niveles de HDL en pacientes con lupus e individuos sanos, no se obtienen resultados concluyentes, pues algunos estudios los encuentran disminuidos, mientras otras investigaciones indican que los niveles son iguales en pacientes con lupus y la población general.

Objetivo: comparar los niveles de colesterol HDL en pacientes con lupus eritematoso sistémico e individuos sanos, según estudios reportados en la literatura científica mundial

Métodos: se realizó una revisión sistemática de la literatura con metanálisis. La búsqueda de los estudios previos se hizo en 6 bases de datos y un ejemplo de la estrategia de búsqueda utilizada fue `systemic lupus erythematosus [Title/Abstract] AND (High density lipoprotein [Title/Abstract] OR HDL [Title/Abstract])`.

Resultados: Se incluyeron 9 estudios en la síntesis cuantitativa. Para hacer el metanálisis, es necesario que los estudios previos reporten la media, junto con la desviación estándar, de los niveles de HDL, tanto en el grupo de pacientes con lupus como en los individuos sanos. Con esa información, se construyó la Tabla 12.

Tabla 12
Descripción de los niveles de colesterol HDL reportada en cada estudio

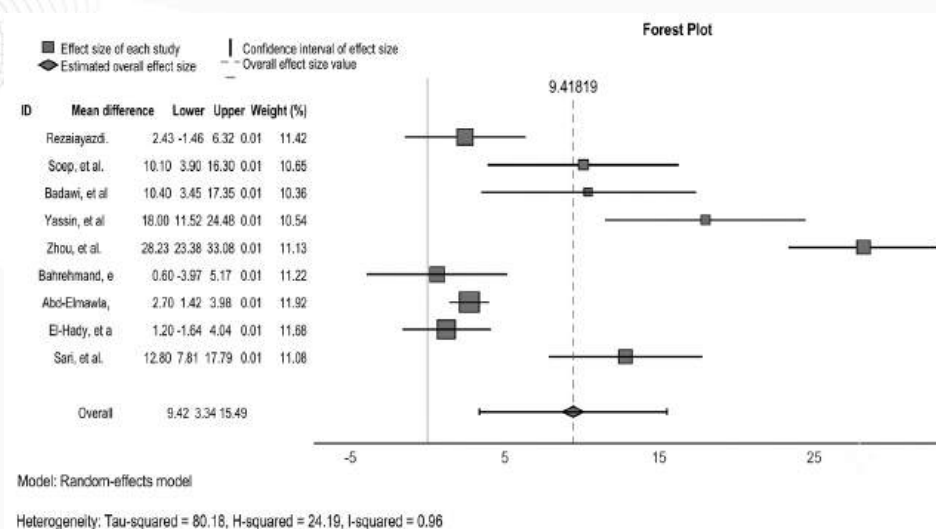
Autor	Tamaño de la muestra sanos	Media del cHDL en sanos	Desviación estándar cHDL sanos	Tamaño de la muestra pacientes con lupus	Media del cHDL en pacientes con lupus	Desviación estándar cHDL en pacientes con lupus
Autor 1	31	47,1	8,2	59	44,7	9,3
Autor 2.	30	50,7	12	33	40,6	13
Autor 3	20	49,7	6,5	30	39,3	14,9
Autor 4	29	61	6	38	43	17
Autor 5	71	60,3	16,3	71	32,1	13,1
Autor 6	103	42,7	11	109	42,1	21,1
Autor 7	35	49,1	2,5	65	46,4	3,4
Autor 8	25	51,5	7,3	50	50,3	5,1
Autor 9	26	46	8,3	24	33,2	9,7

Nota. cHDL: Colesterol de lipoproteínas de alta densidad. Elaboración propia.

En el forest plot se observa una alta heterogeneidad en los estudios individuales, por lo que se realiza un análisis bajo el modelo de efectos aleatorios¹³. En tres de los estudios incluidos no se encontraron diferencias significativas en los niveles de HDL de los pacientes con lupus, en comparación con la población general; sin embargo, seis de los estudios mostraron una disminución significativa en los niveles de HDL en los pacientes con lupus. Coherente con lo anterior, al combinar todos los estudios se encontró que los pacientes con lupus tienen menores niveles de colesterol HDL con una diferencia de medias de -9,42 (IC 95 % -15,4 a -3,34) (ver Figura 20).

Figura 20
Forest plot para la diferencia de medias en los niveles de HDL

¹³ Es importante aclarar que el forest plot, en sentido estricto, no permite interpretar la heterogeneidad; sin embargo, en la mayoría de los programas para hacer metanálisis este gráfico se acompaña de los estadísticos de heterogeneidad, como se observa en la parte inferior que reporta I² de 0,96 y la prueba de hipótesis según la cual se rechaza la hipótesis nula que plantea homogeneidad (valor p = 0,00).



Nota. Elaboración propia.

Conclusión: este metanálisis de estudios transversales analíticos puso al descubierto que los pacientes con lupus tienen niveles de HDL significativamente menores a la población general.

Referencias

- Amare, Y., Gelgalo, D., Pozsgai, É., & Kiss, I. (2025). Systematic Review and Meta-Analysis of Human Papillomavirus Prevalence and Genotypic Disparities Among Human Immunodeficiency Virus-Positive Women in Africa [Revisión sistemática y metanálisis de la prevalencia del virus del papiloma humano y las disparidades genotípicas entre mujeres con virus de inmunodeficiencia humana (VIH) en África]. *Journal of Clinical Medicine*, 14(17), Article 5924. <https://doi.org/10.3390/jcm14175924>
- Aschengrau, A., & Seage G.R. (2020). *Essentials of Epidemiology in Public Health* [Fundamentos de epidemiología en salud pública]. Jones & Bartlett Learning.
- Cardona-Arias, J. A., Higueta-Gutiérrez L. F., & Ríos Osorio, L. A. (2016). *Revisiones sistemáticas de la literatura científica: la investigación teórica como principio para el desarrollo de la ciencia básica y aplicada*. Universidad Cooperativa de Colombia.

- Carvajal Carvajal, C. (2017). Síndrome metabólico: definiciones, epidemiología, etiología, componentes y tratamiento. *Medicina Legal de Costa Rica*, 34(1), 175-193. <https://www.scielo.sa.cr/pdf/mlcr/v34n1/2215-5287-mlcr-34-01-175.pdf>
- Celentano, D. D., Szklo, M., & Farag Y. M. K. (2024). *Gordis. Epidemiología* (7 Ed.). Elsevier.
- Debnath, A., Goel, K., P, A., Shamim, M. A., Satapathy, P., & Gandhi, A. P. (2025). Workplace sexual harassment and violence among women: a systematic review and meta-analysis [Acoso sexual y violencia entre mujeres en el lugar de trabajo: una revisión sistemática y un metanálisis]. *Women & Health*, 65(4), 287-301. <https://doi.org/10.1080/03630242.2025.2478378>
- Downes, M. J., Brennan, M. L., Williams, H. C., & Dean, R. S. (2016). Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS) [Desarrollo de una herramienta de evaluación crítica para evaluar la calidad de estudios transversales (AXIS).]. *BMJ Open*, 6(12), Article e011458. <https://doi.org/10.1136/bmjopen-2016-011458>
- Haby, M. M., Pinart, M., Elias, V., & Reveiz, L. (2018). Prevalence of asymptomatic Zika virus infection: a systematic review [Prevalencia de la infección asintomática por el virus del Zika: una revisión sistemática]. *Bulletin of the World Health Organization*, 96(6), 402-413D. <https://doi.org/10.2471/BLT.17.201541>
- JBI. (2020). Checklist for analytical cross sectional studies [Lista de verificación para estudios transversales analíticos]. https://jbi.global/sites/default/files/2020-08/Checklist_for_Analytical_Cross_Sectional_Studies.pdf
- Li, H., Mu, M., He, Y., Wang, J., Cai, Z., Tang, H., Zhang, B., Luo, H., & Zeng, W. (2025). Prevalence of workplace bullying among medical students: A meta-analysis and systematic review protocol [Prevalencia del acoso laboral entre estudiantes de medicina: un metanálisis y un protocolo de revisión sistemática]. *PloS one*, 20(2), Article e0310076. <https://doi.org/10.1371/journal.pone.0310076>
- Ma, L., Wang, Y., Yang, Z., Huang, D., Weng, H., & Zeng, X. (2020). Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? [Herramientas de evaluación de la calidad metodológica (riesgo de sesgo) para estudios médicos primarios y secundarios:

¿qué son y cuál es mejor?]. *Military Medical Research*, 7(7). <https://doi.org/10.1186/s40779-020-00238-8>

Moreira, T. R., Lemos, A. C., Colodette, R. M., Gomes, A. P., & Batista, R. S. (2019). Prevalence of tuberculosis in incarcerated populations: systematic review and meta-analysis [Prevalencia de tuberculosis en poblaciones carcelarias: revisión sistemática y metanálisis]. *Pan American Journal of Public Health*, 43, Article e16. <https://doi.org/10.26633/RPSP.2019.16>

Munn, Z., Moola, S., Lisy, K., Riitano, D., & Tufanaru, C. (2017). Chapter 5: Systematic reviews of prevalence and incidence [Capítulo 5: Revisiones sistemáticas de prevalencia e incidencia]. In E. Aromataris, & Z. Munn (Eds.). *Joanna Briggs Institute Reviewer's Manual*. The Joanna Briggs Institute. <https://reviewersmanual.joannabriggs.org/>

Netyam, R., Shashank, S., Saravanan, C. R., Kolanu, N. D., Karra, N., Narla, S. J. S. R., Prajjwal, P., Marsool, M. D. M., & Hussin, O. A. (2025). Poor sleep quality (PSQ) prevalence in patients with functional dyspepsia (FD): a systematic review and meta-analysis [Prevalencia de mala calidad del sueño (PSQ) en pacientes con dispepsia funcional (DF): una revisión sistemática y metanálisis]. *Annals of Medicine and Surgery*, 87(9), 5936-5944. <https://doi.org/10.1097/MS9.00000000000003616>

Pelletti, G., Boscolo-Berto, R., Anniballi, L., Giorgetti, A., Pirani, F., Cavallaro, M., Giorgini, L., Fais, P., Pascali, J. P., & Pelotti, S. (2024). Prevalence of alcohol-impaired driving: a systematic review with a gender-driven approach and meta-analysis of gender differences [Prevalencia de la conducción bajo los efectos del alcohol: una revisión sistemática con un enfoque de género y un metanálisis de las diferencias de género]. *International Journal of Legal Medicine*, 138(6), 2523-2540. <https://doi.org/10.1007/s00414-024-03291-3>

Puthran, R., Zhang, M. W., Tam, W. W., & Ho, R. C. (2016). Prevalence of depression amongst medical students: a meta-analysis [Prevalencia de depresión entre estudiantes de medicina: un metanálisis]. *Medical Education*, 50(4), 456-468. <https://doi.org/10.1111/medu.12962>

Rothman, K. J., Greenland, S., & Lash T. L. (2008). *Modern Epidemiology* [Epidemiología moderna]. Lippincott Williams & Wilkins.

Salari, N., Asgharpour, M., Daneshkhah, A., Zarei, H., Asgharpour,

- B., Faghihi, S. H., & Mohammadi, M. (2025). Global prevalence of metabolic syndrome among individuals with bipolar disorder: A systematic review and meta-analysis [Prevalencia global del síndrome metabólico entre personas con trastorno bipolar: una revisión sistemática y metanálisis]. *Archives of Psychiatric Nursing*, 58, Article 151942. <https://doi.org/10.1016/j.apnu.2025.151942>
- Schlosser, R. W., Wendt, O., Bhavnani, S., & Nail-Chiwetalu, B. (2006). Use of information-seeking strategies for developing systematic reviews and engaging in evidence-based practice: the application of traditional and comprehensive Pearl Growing. A review [Uso de estrategias de búsqueda de información para desarrollar revisiones sistemáticas y aplicar prácticas basadas en la evidencia: la aplicación del cultivo de perlas tradicional e integral. Una revisión]. *International Journal of Language & Communication Disorders*, 41(5), 567-582. <https://doi.org/10.1080/13682820600742190>
- STROBE. (2025). *What is STROBE?* [¿Qué es STROBE?]. <https://www.strobe-statement.org/>
- Szklo, M., & Nieto, F.J. (2014). *Epidemiology. Beyond the basics* [Epidemiología. Más allá de lo básico] (3 Ed.). Jones & Bartlett Learning.
- Vaziri, S., Fakouri, F., Mirzaei, M., Afsharian, M., Azizi, M., & Arab-Zozani, M. (2019). Prevalence of medical errors in Iran: a systematic review and meta-analysis [Prevalencia de errores médicos en Irán: una revisión sistemática y un metanálisis]. *BMC Health Services Research*, 19(1), Article 622. <https://doi.org/10.1186/s12913-019-4464-8>

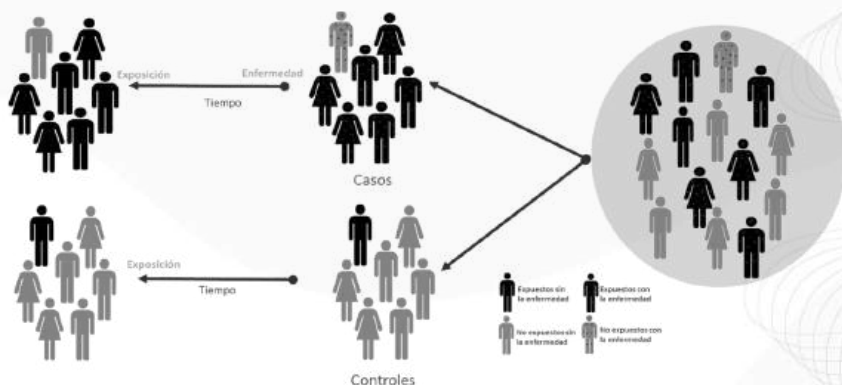
Capítulo 3. Etiología de la enfermedad: síntesis de la literatura de estudios de casos y controles

1. Características de los estudios de casos y controles

Los estudios de casos y controles son un tipo de estudio observacional analítico que tiene como sello distintivo el hecho de que los sujetos se seleccionan en función de si tienen o no la enfermedad. Los sujetos que tienen la enfermedad se denominan casos y los que no la presentan serían los controles; para ambos grupos se buscan registros previos sobre la exposición a posibles factores de riesgo, y se compara la magnitud de la exposición entre casos y controles para identificar si existen diferencias que pueden haber llevado a la enfermedad. Las exposiciones pueden ser factores biológicos, comportamentales, condiciones de vida, condiciones ambientales o tratamientos que pueden modificar el riesgo de enfermar. En este sentido, algunos autores sostienen que los estudios de casos y controles se mueven del efecto a la causa (Aschengrau & Seage, 2020; Celentano et al., 2024) (ver Figura 21).

Figura 21

Representación gráfica de los estudios de casos y controles



Nota. Elaboración propia.

Este tipo de estudios se usa como primer paso cuando se busca establecer la causa o los factores de riesgo para una enfermedad. Usualmente, en las primeras fases de la investigación sobre la etiología de las enfermedades se sospecha de una amplia variabilidad de exposiciones que pueden condicionar el desarrollo del evento, pero no se cuenta con evidencia para sugerir una asociación de cualquiera de esas exposiciones que pueda orientar intervenciones o estrategias de prevención. En esas ocasiones, se puede generar la evidencia usando el diseño de casos y controles, de manera tal que si la proporción de la exposición a ciertos factores entre los casos es mayor que la proporción de esa misma exposición entre los controles se concluiría que dicha exposición es un posible factor de riesgo para la enfermedad¹⁴ (Celentano et al., 2024).

Cuando nos referimos a factores de riesgo hacemos alusión a los atributos o comportamientos que aumentan la probabilidad de que una exposición específica dé lugar a un resultado negativo para la salud. Así, los factores de riesgo se clasifican en modificables cuando incluyen aspectos como la malnutrición, el tabaquismo o el sedentarismo; y no modificables cuando se refieren a características como la edad, antecedentes familiares de la enfermedad o mutaciones en el genoma (Moola et al., 2017). Además, con el diseño de casos y controles también se pueden identificar posibles factores protectores; es decir, aquellas características o exposiciones asociadas con una menor probabilidad de ocurrencia de la enfermedad. Por ejemplo, al llevar una alimentación saludable o practicar regularmente ejercicio físico, se reduce la probabilidad de enfermedades cardiovasculares¹⁵.

El primer paso en el diseño de este tipo de estudios consiste en identificar un grupo de individuos con la enfermedad de interés (los casos) y, para fines de comparación, un grupo de personas sin esa enfermedad (los controles). Luego se evalúan exposiciones previas mediante encuestas, revisión de historias clínicas o cuantificación de analitos en líquidos corporales o tejidos. En términos generales esas exposiciones son dicotómicas; es decir, hay dos posibilidades de respuesta: si el individuo estuvo o no estuvo expuesto al factor

¹⁴ Si bien se tiende a pensar que este proceso estadístico (comparar la exposición entre casos y controles) es sinónimo de identificar factores de riesgo, realmente es más preciso indicar que este proceso analítico es el primer paso para identificar los factores de riesgo, dado que la definición de relación etiológica (factor de riesgo) implica no solo cumplir este criterio que corresponde a la fuerza de asociación, sino que también debe demostrarse la consistencia, especificidad, temporalidad, gradiente biológico, plausibilidad biológica, coherencia, analogía, etc.

¹⁵ La epidemiología de los factores de riesgo ha sido objeto de múltiples críticas por el hecho de considerarla lineal (en la relación causa efecto), reduccionista (en las ontologías de la persona y de la enfermedad), predominantemente biomédica (o circunscrita a la dimensión biológica), que no tiene presente los elementos sociales, culturales e históricos de la enfermedad. Pese a las críticas, este paradigma del riesgo sigue siendo altamente efectivo para conocer e intervenir las principales causas de múltiples enfermedades en general y, en el caso particular de los estudios de casos y controles, para las enfermedades raras o de baja frecuencia.

de interés. Con esta información se divide la población en cuatro grupos: a) los casos que estuvieron expuestos, b) los controles que estuvieron expuestos, c) los casos que no estuvieron expuestos, y d) los controles que no estuvieron expuestos (ver Tabla 13).

Tabla 13
Posibles grupos en un estudio de casos y controles

Tabla de 2X2 en un estudio de casos y controles		
	Casos	Controles
Expuestos	a	b
No expuestos	c	d

Nota. Elaboración propia.

Con base en esos cuatro grupos, los investigadores calculan un número llamado odds ratio (OR) o razones de odds. La OR generalmente se calcula con base en los productos cruzados, es decir: $(a \times d) / (b \times c)$; en este caso, dichos productos solo se deben tomar para el cálculo matemático, pero no es correcto intentar interpretarlos con referencia a lo indicado en la tabla 13 (no tiene sentido interpretar de forma literal la multiplicación de casos expuestos -a- con los controles no expuestos -d-).

Una odds es la probabilidad del evento dividida por su complemento (la probabilidad de no evento). Por su parte, la razón de odds de la enfermedad sería la división de dos odds:

- a) Probabilidad de cumplir un evento –ser caso- $(a/a+b)$ / la probabilidad de no cumplirlo –ser control- $(b/a+b)$ EN LOS EXPUESTOS.
- b) Probabilidad de cumplir un evento –ser caso- $(c/c+d)$ / la probabilidad de no cumplirlo –ser control- $(d/c+d)$ EN LOS NO EXPUESTOS.

La razón de odds de la exposición sería:

- a) Probabilidad de cumplir un evento –ser expuesto- $(a/a+c)$ / la probabilidad de no cumplirlo –ser NO expuesto- $(c/a+c)$ EN LOS CASOS.
- b) Probabilidad de cumplir un evento –ser expuesto- $(b/b+d)$ / la probabilidad de no cumplirlo –ser NO expuesto- $(d/b+d)$ EN LOS CONTROLES

Al despejar cualquiera de las dos razones da $(a \times d / b \times c)$, de lo cual se concluye que las OR de la enfermedad son iguales a las

razones de la exposición.

Supongamos que se hizo un estudio de casos y controles para analizar la relación entre la resistencia a ertapenem (Un grupo de antibióticos) y hospitalización previa en pacientes con infección urinaria.

Tabla 14

Datos para calcular la odds ratio de la hospitalización

	Hospitalizado	No hospitalizado
Infección resistente a ertapenem	80	20
Infección sensible a ertapenem	40	60

Nota. Elaboración propia.

Para la OR de la hospitalización mediante productos cruzados tenemos:

$$OR = (80 \times 60) / (20 \times 40) = 4800/800 = 6$$

Mediante fracciones tenemos lo siguiente:

$$\text{Odds de hospitalización con infección resistente} = \frac{(80/(80+20))}{(20/(80+20))}$$

$$\text{Odds de hospitalización con infección sensible} = \frac{(40/(40+60))}{(60/(40+60))}$$

$$OR = (80 \times 60)/(20 \times 40) = 4800/800 = 6$$

Para la odds Ratio de la resistencia tenemos (ver Tabla 15).

Tabla 15

Datos para calcular la odds ratio de la resistencia

	Infección resistente a ertapenem	Infección sensible a ertapenem
Hospitalizado	80	40
No hospitalizado	20	60

Nota. Elaboración propia.

$$\text{Odds de resistencia en paciente hospitalizado} = \frac{(80/(80+40))}{(40/(80+40))}$$

$$\text{Odds de resistencia en paciente No hospitalizado} = \frac{(20/(20+60))}{(60/(20+60))}$$

$$OR = (80 \times 60)/(40 \times 20) = 4800/800 = 6$$

Las razones de odds se han traducido al español como razón de momios, razón de ventajas, razón de productos cruzados, oportunidad relativa y razón de chances; sin embargo, ninguno de estos términos refleja correctamente el concepto de odds (Cerdeja et al., 2013). Aunque no existe una traducción para el término odds, el concepto más aceptado en algunos libros de epidemiología es el de *posibilidades* (Celentano et al., 2024); la odds ratio (OR) es la medida epidemiológica típica en los estudios de casos y controles (Aschengrau & Seage, 2020; Celentano et al., 2024).

La OR puede tomar valores entre cero e infinito, el 1 es el valor nulo que indica que la exposición es igual en los casos y los controles, los valores mayores a 1 indican que las personas expuestas tuvieron más posibilidades de desarrollar la enfermedad que las personas no expuestas (por lo que comúnmente se interpreta como factor de riesgo) y los valores menores a 1 indican que las personas expuestas tuvieron menos posibilidades de desarrollar la enfermedad que las no expuestas (por lo que comúnmente se interpreta como factor de protección) (ver Figura 22). No obstante, no todo lo que surge de una OR puede catalogarse categóricamente como riesgo o protección, porque en ocasiones simplemente se está cuantificando cuántas veces se presenta un evento en un grupo con respecto a otro. La medida de asociación es una característica, pero el **riesgo no es reducible a** la medida de asociación porque hay criterios de causalidad que se deben evaluar, entre ellos, los de Bradford Hill. Adicional a ello, la OR podría ser producto de una confusión residual cuando se desconoce si se está ajustando por las variables relevantes.

Figura 22

Representación gráfica de los posibles valores que puede tomar la Odds ratio



Nota. Elaboración propia.

Es necesario tener en cuenta que es más importante definir si una exposición es de riesgo o de protección, dado que la odds ratio en algunos casos solo refleja cual grupo tiene mayor frecuencia del evento.

Para decidir en qué situaciones es apropiado desarrollar un estudio de casos y controles se tienen en cuenta cuatro aspectos fundamentales. Primero, **un estudio de casos y controles es particularmente apropiado cuando se estudian enfermedades de baja frecuencia (20 % o menos), porque así se evita el problema del “abandono” o “pérdida de seguimiento” de los participantes** presentes en los estudios de cohorte (Aschengrau & Seage, 2020; Celentano et al., 2024; Moola et al., 2017).

En segundo lugar, un estudio de casos y controles es preferible cuando la enfermedad tiene largos períodos de incubación; es decir, cuando transcurre mucho tiempo entre la exposición y el desarrollo de la enfermedad. Esto debido a que, por su carácter retrospectivo, la exposición y la enfermedad ya ocurrieron al momento de la investigación; mientras que en un diseño de cohorte sería necesario muchos años de seguimiento, un alto presupuesto y una logística compleja (Aschengrau & Seage, 2020; Celentano et al., 2024).

Tercero, los estudios de casos y controles son preferibles cuando se sabe poco sobre la enfermedad de interés, porque permiten a los investigadores evaluar múltiples hipótesis sobre los factores de riesgo (Aschengrau & Seage, 2020; Celentano et al., 2024). Finalmente, esta modalidad de investigación es apropiada cuando se tiene poco presupuesto y se necesita información rápidamente, ya que es posible identificar casos para estudio a partir de historias clínicas o registros hospitalarios, y en ellos evaluar tanto la exposición como la enfermedad (Celentano et al., 2024). En la Tabla 16 se resumen las principales características de los estudios de casos y controles.

Tabla 16
Características distintivas de un estudio de casos y controles

Característica	Estudio de Casos y Controles
Objetivo principal	Comparar la exposición a un factor (que potencialmente es de riesgo) entre personas con la enfermedad (casos) y personas sin la enfermedad (controles).

Tipo de estudio	Observacional, retrospectivo.
Dirección del estudio	Se inicia con la identificación de los casos (enfermos) y los controles (no enfermos), y se retrocede en el tiempo para analizar las exposiciones. Se requiere una fuente secundaria con datos sobre los posibles factores de riesgos; es decir, es longitudinal retrospectivo porque la medición de las variables dependientes se da en un momento del tiempo que es diferente al de la medición de las variables independientes.
Selección de participantes	Se seleccionan dos grupos: casos (individuos con la enfermedad) y controles (individuos sin la enfermedad).
Muestra	Generalmente la muestra es pequeña.
Medición de la exposición	La exposición a factores de riesgo se mide retrospectivamente, lo que puede introducir sesgos (ej. sesgo de recuerdo).
Costo	Menor costo en comparación con los estudios prospectivos, ya que se utiliza información de exposiciones pasadas.
Tiempo requerido	Relativamente rápido, ya que se analiza información de eventos que ya han ocurrido.
Principal medida de asociación	Odds ratio (OR), que estima la relación entre la exposición y la enfermedad.
Ventajas	Útil para enfermedades raras y larga incubación; relativamente rápido y económico.
Desventajas	Mayor susceptibilidad a sesgos; no permite calcular incidencia ni calcular prevalencia.

Nota. Elaboración propia.

Es importante tener en cuenta que una práctica común en un grupo de investigadores es forzar los estudios descriptivos para que parezcan de diseños de casos y controles. Por ejemplo, al estudiar la prevalencia de diabetes, los investigadores toman como casos todos los que tienen diabetes y como controles al subgrupo de los participantes sin diabetes; sin embargo, al no tener el elemento retrospectivo ni seleccionar el grupo apropiado de controles, los resultados estarán sesgados. Esta situación resalta la importancia de contar con estudios previos que generen las principales hipótesis a probar en el diseño de casos y controles.

2. Variantes de los estudios de casos y controles

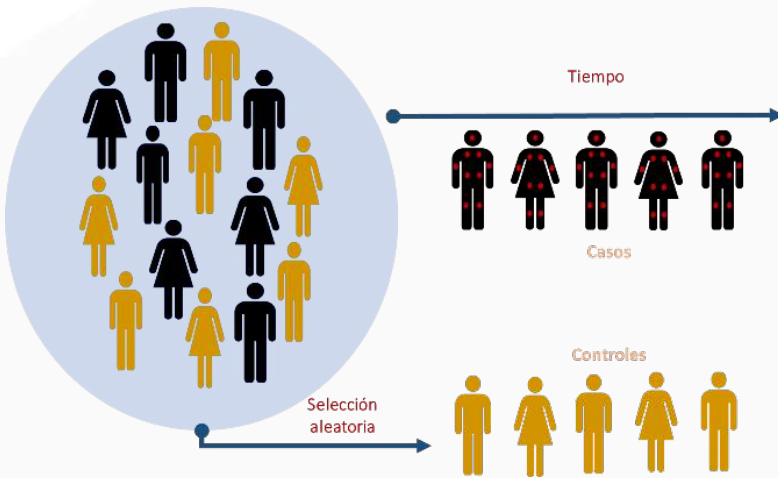
En los últimos años se han propuesto diferentes variantes al diseño clásico de los estudios de casos y controles, las más comunes son:

Estudios casos y controles anidado a una cohorte: En este dise-

ño, la definición de casos y controles está anidada a un estudio de cohorte. Para desarrollar esta variante es necesario: a) una cohorte a la que se le realizará el seguimiento, b) los casos serán los individuos que desarrollen la enfermedad en la medida en la que el seguimiento progrese, y c) los controles se seleccionarán de los individuos de la cohorte que no han desarrollado la enfermedad durante el seguimiento (Lazcano-Ponce et al., 2001) (ver Figura 23).

Figura 23

Representación gráfica del diseño de caso-cohorte

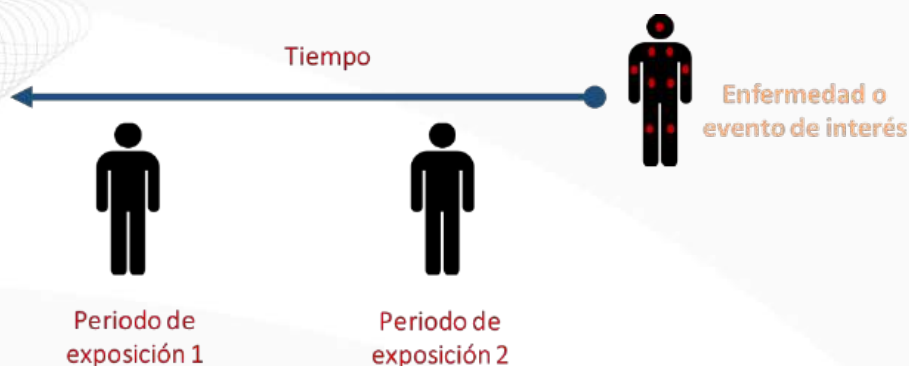


Nota. Elaboración propia.

Estudio caso-caso: En esta variante, se hace la comparación entre el mismo individuo en dos momentos. En el primer momento se miden las exposiciones del paciente a determinados factores mientras no desarrolla la enfermedad. Por su parte, en el segundo momento el paciente se expuso al factor de riesgo y desarrolló la enfermedad. En este diseño cada individuo actúa como su propio control (Lazcano-Ponce et al., 2001) (ver Figura 24).

Figura 24

Representación gráfica del diseño de caso-caso



Nota. Elaboración propia.

3. Validez interna y externa en los estudios de casos y controles

La calidad metodológica es una preocupación importante en los estudios de casos y controles, por eso es necesario reconocer los aspectos que pueden afectar su validez interna y externa. **Con validez interna nos referimos a los errores cometidos durante el proceso de selección de la muestra de estudio, durante las mediciones o a errores ocasionados por la falta de comparabilidad de los grupos estudiados. Con validez externa hacemos alusión a la capacidad del estudio para generalizar los resultados observados en la muestra hacia toda la población** (Hernández-Avila et al., 2000). En esta modalidad de investigación, los aspectos que podrían afectar la validez se agrupan en el procedimiento para la selección de los casos, la selección de los controles y en la evaluación de la exposición.

- **La selección de los casos**

El primer paso en la selección de casos es justamente establecer la definición del caso; es decir, la combinación de signos y síntomas, exámenes físicos y pruebas diagnósticas que configuran la enfermedad de interés. Esta definición debe ser tan precisa como sea posible, excluir el mayor número de interpretaciones y no dar lugar a ambigüedades. No obstante, si se utilizan criterios muy restrictivos,

podría ser difícil alcanzar el tamaño de la muestra necesaria para la investigación; mientras que, si se utilizan criterios inespecíficos, muchas personas que no tienen la enfermedad de interés se incluirán por error en el estudio (Aschengrau & Seage, 2020). Así que siempre que se tenga la posibilidad de elegir, es preferible inclinarse por una definición restrictiva que por una inespecífica. Por otro lado, también es posible dividir los casos de acuerdo con criterios como la gravedad (leve, moderada, severa); esta clasificación puede ser útil para mejorar la precisión de la definición y reducir errores (Aschengrau & Seage, 2020; Celentano et al., 2024).

El siguiente reto consiste en identificar la fuente de selección de los casos; estas fuentes suelen ser las clínicas o los hospitales (Aschengrau & Seage, 2020; Celentano et al., 2024). En este punto debe tenerse en cuenta el nivel de atención del hospital que se selecciona; es decir, si es un hospital de primero, segundo o tercer nivel, ya que si el hospital del que se extraen los casos es de tercer nivel, por su naturaleza, admitirá selectivamente a un gran número de pacientes gravemente enfermos, así que los factores de riesgo identificados en el estudio pueden ser solo de personas con formas graves de la enfermedad; además, si se seleccionan casos de un solo hospital, los factores de riesgo pueden ser exclusivos de ese hospital y los resultados no pueden generalizarse (Celentano et al., 2024). En consecuencia, si se van a utilizar casos hospitalizados, es conveniente seleccionar los casos de varios hospitales.

Otra consideración importante es si incluyen casos incidentes (casos nuevos) o prevalentes (personas que pueden haber tenido la enfermedad durante algún tiempo). El problema con el uso de casos incidentes es que se debe esperar mucho tiempo a que los casos se diagnostiquen para alcanzar el tamaño de la muestra requerido, mientras que si se usan casos prevalentes los factores de riesgo que se identifiquen pueden estar más relacionados con la supervivencia que con el desarrollo de la enfermedad (incidencia). Esto constituiría un grupo de casos no representativo y los factores de riesgo identificados solo reflejarían las características de los sobrevivientes (Aschengrau & Seage, 2020; Celentano et al., 2024), lo cual dificultaría establecer la temporalidad de la relación (garantizar que la exposición antecede al evento). En este sentido, se prefieren los casos incidentes porque esta modalidad de investigación busca los factores que conducen al desarrollo de la enfermedad en lugar de los factores que afectan la duración de la enfermedad (Aschengrau & Seage, 2020; Celentano et al., 2024).

- **Selección de controles**

La forma en que se seleccionan los controles es determinante para poder concluir que existe una asociación entre la exposición y la enfermedad (Celentano et al., 2024). En este sentido, es importante precisar que los controles deben ser una muestra de la población que produjo los casos, de manera que si un participante del grupo control contrae la enfermedad debe terminar en el grupo de los casos. Otro principio importante es que los controles deben muestrearse independientemente del estado de exposición, de modo que los controles expuestos y no expuestos deben tener la misma probabilidad de ser seleccionados para la investigación. Una cuestión conceptual adicional es qué tan similares deben ser los controles a los casos; es decir, si deberían ser similares a los casos en todos los aspectos, excepto en la enfermedad de interés, o en cuáles y en qué aspectos podrían diferir (Celentano et al., 2024).

Para enfrentar estos retos, lo central es tener claro el objetivo del equiparamiento; por ejemplo, si la intención es controlar confusores, entonces los casos y los controles deberían ser iguales en los potenciales confusores de la asociación causa efecto que se está buscando (acorde con las hipótesis del estudio); sin embargo, se recomienda evitar el problema del sobre-equiparamiento que consiste en que no se halla ningún tipo de asociación porque ambos grupos son idénticos en las variables a estudiar.

Las fuentes para la selección de controles generalmente pueden ser individuos de la población general o individuos que asisten a un hospital o clínica (Aschengrau & Seage, 2020; Celentano et al., 2024). Incluir controles de población general implica mayores costos y una logística más compleja; además, al tratarse de individuos generalmente saludables, su recuerdo de exposiciones pasadas puede ser menos preciso que el de los casos. Por su parte, los pacientes hospitalizados a menudo se seleccionan como controles debido a que son “población cautiva” y de fácil acceso; sin embargo, no representan la población de referencia porque al estar hospitalizados difieren de las personas en la comunidad y, por tanto, difieren en los niveles de exposición, así que no se pueden generalizar los resultados (Celentano et al., 2024).

Por otra parte, existe una dificultad relacionada con la elección de las enfermedades que deben tener los controles para su inclusión. Pese a las limitaciones de la inclusión de controles hospitalarios, se destacan que son fáciles de identificar, tienen buenas tasas de parti-

cipación, son menos costosos de identificar que los de población general y su recuerdo de exposiciones pasadas será similar a la de los casos, porque también están enfermos (Aschengrau & Seage, 2020; Celentano et al., 2024).

Una cuestión adicional en la selección de los controles es la elección de cuántos se deben incluir por cada caso. En este punto se recomienda que cuando el número de casos disponibles es limitado y los controles son abundantes, es posible aumentar la potencia estadística del estudio, aumentando el tamaño del grupo de control; así que podrían incluirse hasta cuatro controles por cada caso (4:1) (Aschengrau & Seage, 2020; Celentano et al., 2024). Sin embargo, lo más usual es 1:1 y 2:1. Por otro lado, es importante aclarar que, debido a que la proporción de controles por caso está determinada por el investigador, claramente no refleja la prevalencia de la enfermedad en la población en la que se realiza el estudio. En el mismo sentido, no es una buena práctica tener más casos que controles con la pretensión de reflejar la situación poblacional, porque se estaría confundiendo el propósito de este tipo de estudios con la investigación descriptiva.

- **La evaluación de la exposición**

Los estudios de casos y controles se utilizan para investigar el riesgo de enfermedad en relación con una amplia variedad de exposiciones; por tanto, los investigadores se esfuerzan por obtener información suficientemente detallada sobre la antecedencia de la exposición, las fuentes, la frecuencia y la duración de estas exposiciones. La precisión es una preocupación particular en los estudios de casos y controles porque los datos de la exposición son retrospectivos. De hecho, las exposiciones relevantes pueden haber ocurrido muchos años antes de la recopilación de datos, lo que dificulta la recolección de información correcta debido a que prácticamente todos estamos limitados en nuestra capacidad para recordar información, podemos mostrarnos reticentes a admitir prácticas que podrían considerarse estigmatizantes o simplemente no tenemos la información solicitada (Celentano et al., 2024).

Una consecuencia derivada de lo anterior es que algunos de los casos o controles que estuvieron expuestos se pueden clasificar de manera errónea como no expuestos, y algunos de los que realmente no estuvieron expuestos se clasificarán erróneamente como expuestos. Esto lleva a subestimar el verdadero riesgo de la enfermedad asociada con la exposición (Celentano et al., 2024). En este sentido,

se recomienda que si no hay buena fuente de datos secundarios y no se garantiza la antecendencia de la exposición, es deseable bajarle el nivel al tipo de estudio y dejar el diseño como analítico transversal.

En cuanto a las limitaciones propias de los estudios de casos y controles se destaca que, debido a la naturaleza retrospectiva de la recopilación de datos, hay una mayor probabilidad de sesgo. Algunos autores han argumentado que, debido a la probabilidad de sesgos, los estudios de casos y controles no son adecuados para detectar asociaciones débiles (con odds ratios menores a 1.5). Además, debido a que la recolección de datos es retrospectiva, puede ser difícil establecer la correcta relación temporal entre la exposición y la enfermedad (Aschengrau & Seage, 2020; Celentano et al., 2024).

Por las limitaciones expuestas, algunos epidemiólogos ortodoxos acuñaron el término despectivo estudio TROHOC (cohorte escrito al revés) para referirse a los estudios de casos y controles (Aschengrau & Seage, 2020; Celentano et al., 2024). Sin embargo, en los últimos años la visión de los epidemiólogos sobre los estudios de casos y controles ha cambiado drásticamente, y ya no se consideran una alternativa inferior a un estudio de cohorte, sino que se considera un diseño altamente eficiente para aprender sobre las relaciones exposición-enfermedad.

4. Indicaciones para sistematizar la evidencia de los estudios de casos y controles y elegir cuándo combinar los resultados

Las revisiones sistemáticas sobre factores de riesgo y etiológicos evalúan la relación entre ciertas exposiciones (modificables o no modificables) y el desarrollo de una enfermedad, afección u otro resultado en salud (Moola et al., 2017). En principio, en términos de salud pública, es posible que se desee conocer toda la gama de efectos que puede tener una exposición en diferentes desenlaces en salud, pero en el contexto de una revisión sistemática es recomendable que se aborde la exposición con cada desenlace de interés por separado (Glasziou et al., 2001).

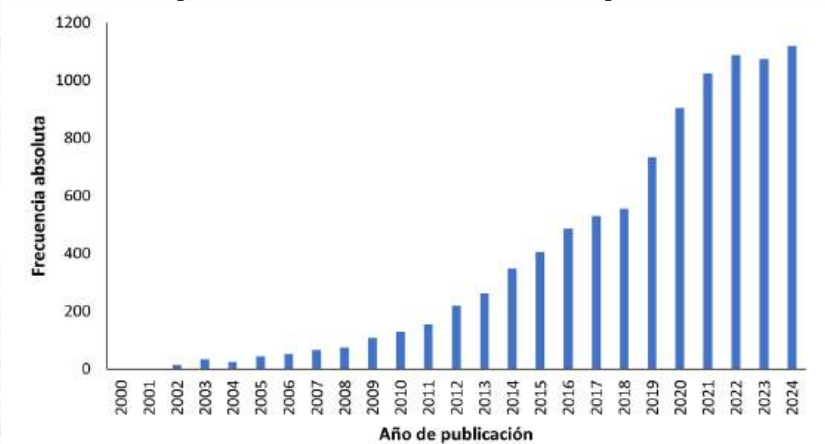
En este contexto, la sistematización de la evidencia relacionada con estudios de casos y controles surge cuando nos enfrentamos a problemas como: i) discrepancias en estudios individuales sobre el hecho de que una exposición constituya un factor de riesgo o un

factor protector para una enfermedad; ii) en los casos en los que la evidencia es consistente en cuanto a que una exposición representa un factor de riesgo para la enfermedad, pero no hay consenso sobre la magnitud del riesgo, la duración o el gradiente de la exposición; iii) las investigaciones con el diseño de factores de riesgo se han realizado en diversas poblaciones, pero todas pequeñas y tienen baja validez externa; y iv) la selección de los casos y los controles en los estudios individuales provienen de clínicas y hospitales, de modo que los factores de riesgo hallados no pueden generalizarse más allá de las muestras estudiadas (Glasziou et al., 2001).

Sumado a los criterios enunciados en el párrafo anterior, es necesario verificar si ya existe una revisión sistemática adecuada sobre el tema de estudio; si no se encuentra dicha revisión o si ninguna coincide directamente con nuestras necesidades, está desactualizada o tiene problemas metodológicos, entonces resulta pertinente emprender la revisión.

En este sentido, se ha realizado una amplia diversidad de revisiones sistemáticas de estudios de casos y controles. En una búsqueda en PubMed, con los términos (Metanalysis[Title]) AND (case-control [Title/Abstract]), se encuentran 7096 artículos, mientras que con la búsqueda (systematic review[Title]) AND (case-control [Title/Abstract]) se encuentran 9466 publicaciones relacionadas con la síntesis de la literatura de estudios de casos y controles entre el 2000 y el 2024 (ver Figura 25).

Figura 25
Frecuencia de publicaciones en PubMed sobre revisiones sistemáticas de estudios de casos y controles entre los años 2000 y 2024



Nota. Elaboración propia.

Esas investigaciones han abordado tópicos como la hemocromatosis y el cáncer de seno (Buttignol et al., 2025), la asociación entre la suplementación con hierro durante el embarazo y el riesgo de leucemia infantil (Dabir et al. 2025), el consumo de lácteos y cáncer de testículo (Signal et al., 2018), el dolor psicológico con la ideación suicida y suicidio (Ducasse, Holden, Boyer, et al., 2018), entre otros. Al igual que en la síntesis de la literatura sobre estudios de prevalencia, este método se ha aplicado a tópicos relacionados con los servicios de salud que no necesariamente se asocian con una enfermedad; por ejemplo, la exposición a violencia de pareja durante el embarazo y el desarrollo de problemas de conducta en niños y adolescentes (Silva et al., 2018), desempleo de los padres y experiencias adversas en la infancia (Judd et al. , 2023), apnea obstructiva del sueño y riesgo de accidentes automovilísticos (Tregear et al., 2009), la exposición a triclosán y la disminución en la calidad espermática (Adegbola et al., 2024), entre otros.

4.1 Características del protocolo de búsqueda y selección de estudios en la síntesis de estudios transversales

Las revisiones sistemáticas de la literatura sobre esta modalidad de investigación deben seguir los mismos principios básicos de las revisiones de otros tipos de estudios; es decir, deben contar con un protocolo a priori en el que se especifique la estrategia de búsqueda, los criterios de inclusión y exclusión, así como la extracción y el análisis de la información. De manera particular, deben describir la exposición, la población o grupos en riesgo y la enfermedad, síntoma o resultado de salud de interés. En algunos casos puede ser importante incluir el contexto o la delimitación geográfica específica de los estudios que serán objeto de la síntesis (Glasziou et al., 2001).

4.1.1 La estrategia de búsqueda

La estrategia de búsqueda deberá estar alineada con la pregunta que orienta la investigación. Quienes emprendan una revisión sistemática de estudios sobre factores de riesgo, rápidamente se darán cuenta de que el formato PICO tradicional (Population, Intervention, Control, Outcome) es coherente con las revisiones sistemáticas que evalúan la efectividad de las intervenciones o terapias médicas,

pero no se alinean con las preguntas relacionadas con factores etiológicos o de riesgo. En este sentido, han emergido algunas modificaciones a la estrategia para la formulación de la pregunta, entre ellas, la propuesta por Dekkers y colaboradores sintetizada como PECO (Population, Exposure, Control, Outcome) (Dekkers et al., 2019; Morgan et al., 2018) o la propuesta por el instituto Joanna Briggs conocida como PEO (Population, Exposure, Outcome) (Moola et al., 2017). Cualquiera que sea la estrategia elegida, en el algoritmo de búsqueda se debe especificar la población (tipos de participantes), la exposición de interés (variable independiente) y la enfermedad de interés (variable dependiente).

Sumado a esto, en los casos en los que la información es prolífica es conveniente incluir en la estrategia el tipo de estudio (casos y controles). Además, delimitar el rango de años buscados y cualquier restricción de idioma (por ejemplo, solo se considerarán para inclusión los estudios publicados en inglés y español). Por otro lado, si no se conoce mucho del campo, basta con elaborar una estrategia de búsqueda que incluya la enfermedad y el tipo de estudio (casos y controles); de esta forma se abordaría, desde un enfoque amplio, la evidencia disponible sobre asociaciones etiológicas.

Como se mencionó en el capítulo anterior, para la enfermedad, problema de salud de interés o, en este caso, la exposición, se deben utilizar sinónimos que pueden o no estar en las plataformas de lenguaje controlado como los términos Decs y/o Mesh. De este modo, una revisión sistemática que quiera dar cuenta del riesgo que representan los polimorfismos en el gen XRCC3 para el desarrollo de cáncer de seno, deberá tener en cuenta tanto los sinónimos de la exposición “X-ray repair cross-complementing group 3” OR XRCC3; como los sinónimos del desenlace “Cáncer de seno” OR “Carcinoma mamario” OR “Neoplasia de seno” OR “Neoplasia mamaria” OR “Tumor de seno” OR “Tumor de mama”.

La elección de las palabras clave, su combinación y la restricción a algunas partes del texto (título/resumen) dependerá de si el investigador prefiere una estrategia sensible o específica. Como ya vimos, una estrategia de búsqueda muy sensible no requiere restricciones en la búsqueda y, por tanto, recuperará una gran cantidad de información que puede no estar relacionada directamente con el tema; mientras que una estrategia específica o un poco más restrictiva, con palabras limitadas al título o al resumen, asegura la recuperación de una serie de estudios realmente vinculados con el tópico de interés (Glasziou et al. , 2001). Finalmente, es fundamental especificar las

plataformas en las que hará la búsqueda, pues de su número y pertinencia dependerá la exhaustividad de la revisión. En la Tabla 17 se presenta el ejemplo de la construcción de la estrategia de búsqueda para el estudio propuesto.

Tabla 17
Frecuencia absoluta de los artículos identificados en cada base de datos y con cada estrategia de búsqueda

PubMed Title/Abstract		
AND	XRCC3	"x-ray repair cross-complementing group 3"
"Cáncer de seno"	n	N
"Carcinoma mamario"	n	N
"Neoplasia de seno"	n	N
"Neoplasia mamaria"	n	N
"Tumor de seno"	n	N
"Tumor de mama"	n	N
Scielo Title/Abstract		
AND	XRCC3	"x-ray repair cross-complementing group 3"
"Cáncer de seno"	n	N
"Carcinoma mamario"	n	N
"Neoplasia de seno"	n	N
"Neoplasia mamaria"	n	N
"Tumor de seno"	n	N
"Tumor de mama"	n	N
Google Scholar Title		
AND	XRCC3	"x-ray repair cross-complementing group 3"
"Cáncer de seno"	n	N
"Carcinoma mamario"	n	N
"Neoplasia de seno"	n	N
"Neoplasia mamaria"	n	N
"Tumor de seno"	n	N
"Tumor de mama"	n	N

Nota. Elaboración propia.

4.1.2 Los criterios de inclusión

Con los criterios de inclusión se delimitan los artículos objeto de la síntesis; es decir, aquellos sobre los cuales se pueden evaluar la hipótesis o las hipótesis que orientan la búsqueda. En este sentido, algunos criterios de inclusión comunes en la síntesis de la literatura de esta modalidad de investigación son: i) que los artículos recuperados sean estudios realizados en humanos¹⁶, ii) que las investigaciones sigan el diseño de un estudio de casos y controles, y iii) que el artículo se corresponda con los componentes de la pregunta PECO de interés (población, exposición, comparador y desenlace).

Criterios de inclusión más específicos pueden estar relacionados con la población de interés (tipos de participantes), en cuyo caso puede delimitarse según el ciclo vital (pediátricos, adolescentes, adultos, ancianos) o la pertenencia a grupos clave (prisioneros, población LGTBI, migrantes, etc.). La enfermedad o el resultado también se debe informar claramente en esta sección; al hacerlo es importante tener en cuenta que pueden presentarse diferentes criterios en los estudios sobre la manera de definir una misma enfermedad, incluso cuando hay consenso en la definición esta puede cambiar con el tiempo en la medida que se aprende más sobre la afección y sus manifestaciones (Aschengrau & Seage, 2020; Celentano et al., 2024).

4.1.3 Los criterios de exclusión

Los criterios de exclusión se aplican a los estudios que pasaron los criterios de inclusión y su objetivo es prescindir de investigaciones que puedan conducir a errores sistemáticos al momento de analizar la información. En este sentido, es común que en esta sección se incluyan varios criterios: i) algunos relacionados con el uso de técnicas con bajo desempeño diagnóstico, tanto para la definición de la exposición como para el desenlace; ii) que en el estudio no se reporte la medida de asociación epidemiológica (odds ratio) para los factores estudiados, ni los datos que permitan calcularla; iii) que la selección del grupo control no sea apropiada, porque no provienen de la misma población que produjo los casos o sus características no son comparables con ellos; y iv) las investigaciones que reportan los hallazgos de una misma población en dos publicaciones diferentes.

¹⁶ También puede delimitarse a estudios realizados en modelos celulares o en modelos animales.

Aunque este último punto es fenómeno poco común, ocurre y es necesario estar atento a él para evitarlo.

Un punto crucial en esta sección se refiere a la evaluación de las posibles fuentes de confusión; es decir, a los procedimientos que siguieron los autores de los estudios individuales para identificar variables que puedan distorsionar la asociación entre la exposición y la enfermedad de interés y que, por ello, conducen a diferencias importantes entre las estimaciones brutas de una asociación y las ajustadas. En este sentido, debe evaluarse si las publicaciones presentan procedimientos para prevenir la confusión desde el diseño del estudio (criterios de inclusión, emparejamiento) o en el análisis de la información (con estratificación de subgrupos o análisis multivariado) (De Irala et al., 2001).

4.1.4 La extracción de la información

En esta sección se deben incluir los detalles de los datos que se van a extraer de cada estudio. Independientemente del tópico de la revisión, se sugiere que se tenga en cuenta, como mínimo, la siguiente información agrupada en las características propias del estudio, de los participantes y del análisis de la información. En las características del estudio se debe mencionar apellido del autor, año de publicación en la revista, el tipo de estudio, lugar de realización (ciudad, país) y el procedimiento de reclutamiento de los participantes. En las características de los participantes se debe explicitar, tanto de los casos como de los controles, edad, sexo, tamaño de la muestra en cada grupo, la pertenencia algún grupo clave (hospitalizados, prisioneros, LGTBI, etc), exposición de interés (tipo, frecuencia, intensidad, duración) y enfermedad de interés (definición, severidad, métodos de medida).

En el análisis de la información se debe extraer la proporción de casos y controles expuestos y no expuestos, la odds ratio (OR crudas y ajustadas) y las covariables (Moola et al., 2017). Para facilitar posibles metanálisis, estos datos de la OR deberían extraerse en términos de las celdas a, b, c, d para cada exposición evaluada (para las OR crudas), y para las OR ajustadas se debe indicar el modelo estadístico y las variables de ajuste. Adicionalmente, se deberían incluir un grupo de variables con las cuales se pueda realizar meta-regresión; por ejemplo, la edad, el sexo, el país, la calidad de los estudios, entre otros.

Evaluación de la calidad metodológica y riesgos de sesgos

La evaluación de la calidad metodológica o evaluación crítica es un proceso realizado en las revisiones sistemáticas para establecer la validez interna y el riesgo de sesgo de los estudios que cumplen con los criterios de inclusión de la revisión. La evaluación crítica debe ser realizada por dos revisores de manera independiente, y en los casos en los que ellos no están de acuerdo en alguno de los puntos de la evaluación, se puede requerir de un tercero que resuelva las discrepancias. En la evaluación de la calidad metodológica los autores deben hacer explícitos los criterios que van a seguir para determinar si un estudio es de buena, moderada o de mala calidad, y, con base en ello, decidir si van a incluir solo estudios de buena calidad o todos los estudios independientemente de la calidad (Moola et al., 2017).

Pese a la importancia de esta etapa, actualmente no existe una metodología universalmente aceptada para evaluar la calidad metodológica de los estudios de casos y controles. En este sentido, se suele utilizar las listas de chequeo para el reporte de esta modalidad de investigación consignados en la guía STROBE (*STrengthening the Reporting of OBservational studies in Epidemiology*); sin embargo, estas han sido objeto de crítica en algunos escenarios por incluir, además de aspectos metodológicos, algunos elementos editoriales relacionados con las características del título, el resumen o las fuentes de financiación (STROBE, 2025). Una alternativa a la guía STROBE puede encontrarse en la lista de chequeo recomendada por los autores del instituto Joanna Briggs (Moola et al., 2017), el cual incluye 10 ítems que deben valorarse en términos de si se cumplen, no se cumplen, no es claro en el texto o no aplica. Al final, cada revisor debe emitir su concepto en relación con si el artículo se incluye o no en el estudio y argumentar las razones para ello (ver Tabla 18).

Tabla 18

Criterios para evaluar la calidad metodológica de estudios de casos y controles

Criterios para evaluar la calidad metodológica de estudios de casos y controles				
	Cumple	No cumple	No es claro	No aplica
¿Los grupos son comparables además de la presencia de enfermedad en los casos o la ausencia de enfermedad en los controles?				
¿Los casos y los controles están pareados adecuadamente?				
¿Se usaron los mismos criterios para la identificación de los casos y los controles?				
¿La exposición (variable independiente) se midió con técnicas válidas y fiables?				
¿La exposición se midió de la misma manera en los casos y en los controles?				
¿La exposición fue lo suficientemente larga como para ser significativa?				
¿Se identificaron variables de confusión?				
¿Se establecieron estrategias para tratar los factores de confusión?				
¿La enfermedad (variable dependiente) se midió con técnicas válidas y fiables?				
¿El análisis de la información fue apropiado?				

Nota. Elaboración propia.

Sumado a lo anterior, **en los estudios etiológicos o sobre factores de riesgo se debe especificar si los efectos observados en los estudios permiten establecer una asociación causal. Para esto existen varios modelos en medicina que pueden aplicarse para juzgar causalidad, entre ellos, el modelo de Koch-Henle, que es paradigmático en el estudio de enfermedades infecciosas; el modelo de Austin Bradford-Hill, que es aplicable en el estudio de enfermedades no infecciosas; o el modelo de Rothman que contempla relaciones multicausales** (Álvarez-Martínez & Pérez-Campos, 2004).

Entre todos los criterios descritos en estos modelos, es fundamental evaluar los criterios de fuerza de asociación y antecendencia

de la exposición. Para la fuerza de la asociación se debe valorar si el riesgo de desarrollar la enfermedad aumenta cuando aumenta la exposición. Por su parte, la antecedencia se refiere al hecho de establecer una relación temporal entre la exposición (debe ser antes) y la enfermedad.

4.2 Análisis de la información

El análisis de la información de todas las revisiones sistemáticas se presenta en una sección narrativa y una sección numérica. La síntesis narrativa se basa principalmente en el recuento y la descripción de las características de los estudios primarios, el contexto, la calidad y los hallazgos. Generalmente, la descripción parte enunciando la distribución geográfica y temporal de las publicaciones, las cuales se acompañan de tablas, gráficos y otros diagramas que ayudan a complementar la información.

Posteriormente, se presentan características de los participantes (número, edad, sexo), su entorno, la exposición, los tipos de exposición y la enfermedad, junto con cualquier otro factor de interés (por ejemplo, severidad de la enfermedad, pruebas diagnósticas). En algunas ocasiones se pueden hacer subgrupos según los criterios enunciados (por ejemplo, país, tipos y duración de la exposición) y hacer comparaciones a través del cálculo de frecuencias absolutas y relativas junto con los intervalos de confianza del 95 %. En un paso final, previo a la síntesis cuantitativa, se hacen algunos comentarios sobre el diseño y la calidad metodológica de los estudios.

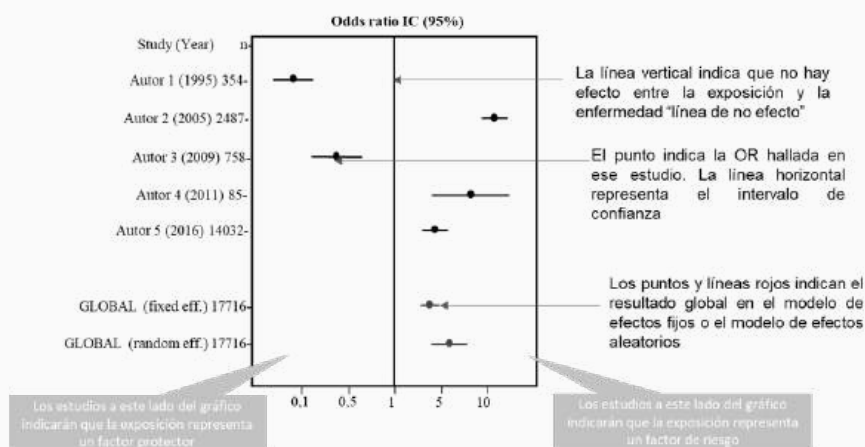
La síntesis numérica se presenta principalmente cuando es posible realizar un metanálisis. Como ya hemos visto, un metanálisis es un procedimiento estadístico que combina los hallazgos de múltiples estudios individuales en una medida global. En el metanálisis se analizan cuatro aspectos fundamentales: i) la heterogeneidad con estadísticos como I^2 (heterogeneidad en valores mayores a 0,5), pruebas de hipótesis como Dersimonian Laird (heterogeneidad en valores $p < 0,05$) o métodos gráficos como Glabraith (heterogeneidad si al menos un estudio está fuera del intervalo); ii) el sesgo de publicación con pruebas de hipótesis como la de Begg (sin sesgos en valores $p \geq 0,05$) o métodos gráficos como el Funnel Plot (sin sesgos cuando los estudios se ubican a ambos lados de la OR estimada); iii) la sensibilidad, generalmente con la tabla o el gráfico de influencias (para confirmar que la OR combinada no cambia al excluir cada uno de

los estudios en fases sucesivas); y iv) la OR combinada o global, y las OR de los estudios individuales, con sus intervalos de confianza.

La interpretación del tamaño del efecto en los metanálisis que abordan cuestiones etiológicas o de factores de riesgo es difícil debido a que las variables incluidas en los análisis multivariados de los estudios individuales pueden ser muy heterogéneas; sin embargo, en la mayoría de las ocasiones es posible hacerlos y el forest plot tiene el aspecto de la Figura 26. Una guía propuesta para la interpretación de los tamaños del efecto sobre factores de riesgo a través del cálculo de odds ratio (OR), sugiere que una OR de 2 es el valor mínimo significativo, un valor de 3 se considera moderadamente significativo y se considera que un valor de 4 indica una gran importancia desde un punto de vista práctico (Moola et al., 2017).

Figura 26

Aspecto de un forest plot para un metanálisis de estudios de casos y controles



Nota. Elaboración propia.

A pesar de la sencillez de las interpretaciones previas, en los estudios incluidos para el metanálisis de la OR, el lector debe tener presente la definición y selección de los casos y los controles, las variables con las cuales realizaron el equiparamiento, la definición y medición de la exposición o posible factor de riesgo, y los demás aspectos indicados al inicio de este capítulo, para poder evaluar críticamente la pertinencia de la OR cruda combinada.

Sumado a lo anterior, es conveniente precisar que en algunas ocasiones los estudios primarios que investigan el riesgo de una ex-

posición presentan los datos en diferentes niveles de exposición o en diferentes categorías que permitan reflejar una relación de “dosis-respuesta” entre la exposición y enfermedad. Cuando los estudios individuales presentan esta información categorizada de manera similar, es posible hacer metanálisis entre esos subgrupos. Finalmente, en algunos entornos se argumenta que no es conveniente combinar resultados cuando se halla heterogeneidad entre los estudios individuales; al respecto, se debe aclarar que la decisión de realizar un metanálisis no debe basarse solo en consideraciones estadísticas con respecto a la heterogeneidad, sino que debe basarse en la pregunta que orienta la revisión, las características de los estudios incluidos y la interpretación o alcance que se le quiera dar a los resultados¹⁷ (Moola et al., 2017).

5. Ejemplo práctico

Problema de investigación: La tuberculosis representa uno de los principales problemas de salud pública en el ámbito mundial. Este problema se ha agravado en los últimos años con el surgimiento de la resistencia a los antibióticos. Estudios previos han tratado de identificar los factores de riesgo para la infección con tuberculosis con resistencia a los medicamentos, y se ha identificado que recibir un esquema de tratamiento previo de antibióticos, de manera inadecuada, podría incrementar el riesgo, pero hay amplia variabilidad en la magnitud del riesgo reportada en los estudios previos (OR fluctúan entre 1,1 y 30,8).

Objetivo: Analizar la relación entre el uso previo de antibióticos y la infección por tuberculosis resistente a este grupo de medicamentos, según estudios reportados en la literatura científica mundial.

Métodos: se realizó una revisión sistemática de la literatura con metanálisis. La búsqueda de los estudios previos se hizo en 5 bases de datos utilizando la estrategia ((Tuberculosis[Title/Abstract]) AND Resistant[Title/Abstract]) AND case control[Title/Abstract].

Resultados: Se incluyeron 8 estudios en la síntesis cuantitativa. De los artículos individuales, se extrajo la información descrita en la Tabla 19, necesaria para realizar el metanálisis.

¹⁷ En este tema es importante aclarar que la heterogeneidad que se evalúa en un metanálisis es estadística, corresponde al grado de variación en el tamaño y precisión del efecto analizado (una prevalencia, una OR, un RR, etc). Esta heterogeneidad estadística no define si se debe realizar o no el metanálisis, lo que determina es el modelo para la estimación de la medida combinada (efectos fijos o aleatorios). Sin embargo, los autores que critican las medidas combinadas con estudios altamente heterogéneos, generalmente no se refieren a estas estimaciones estadísticas, sino a heterogeneidad en los diseños o ejecución de los estudios.

Los casos expuestos representan a los pacientes con infección por tuberculosis resistente que estuvieron previamente sometidos a un esquema de tratamiento con antibióticos; los controles expuestos son los pacientes con infección por tuberculosis con sensibilidad a los antibióticos y que recibieron un esquema previo de tratamiento. Los casos no expuestos representan a los individuos con la infección resistente sin uso previo de antibióticos; y los controles no expuestos son los pacientes con infección por tuberculosis sensible, sin uso previo de antibióticos.

Tabla 19
Información a extraer en un metanálisis de estudios de casos y controles

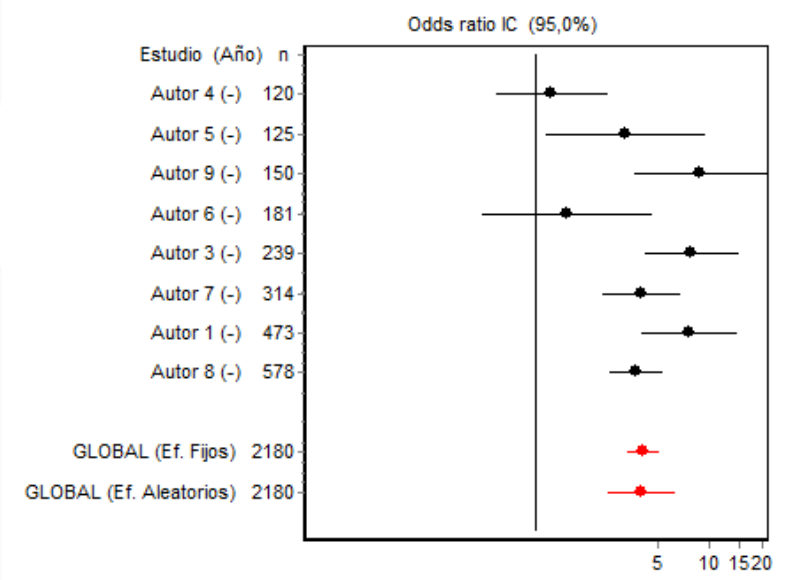
Autor	n total	n ca- sos	n con- troles	n casos expues- tos (a)	N contro- les expues- tos (b)	n casos no ex- puestos (c)	n con- troles no expues- tos (d)
Autor 1	473	64	409	23	28	41	381
Autor 2	239	123	116	73	18	50	98
Autor 3	120	60	60	25	22	35	38
Autor 4	125	50	75	45	55	5	20
Autor 5	181	80	101	7	6	73	95
Autor 6	314	105	209	53	42	52	167
Autor 7	578	287	291	214	127	73	164
Autor 8	150	30	120	20	22	10	98

Nota. Elaboración propia.

Los datos de la Tabla 19 se introducen al software y se calcula la medida combinada. En el forest plot se observa que la OR de los estudios individuales para el uso previo de antibióticos fluctuó entre 1,23 (IC 95 % 0,59-2,57) y 8,90 (IC 95 % 3,66-21,67). En seis de los estudios incluidos la OR muestra que el uso de antibióticos incrementa la posibilidad del riesgo de infección por tuberculosis resistente a los antibióticos y solo dos investigaciones indican que el uso previo de antibióticos no se relaciona con la infección por bacterias resistentes. Coherente con lo anterior, la medida combinada, bajo el modelo de efectos aleatorios, genera un resultado de 4,06 (IC95 % 2,61-6,30), lo que indica que efectivamente el uso previo de antibióticos incrementa la posibilidad de infección por tuberculosis resistente a este grupo de medicamentos y puede considerarse como un factor de riesgo. Es importante aclarar que en este caso se hace

la interpretación del metanálisis bajo el modelo de efectos aleatorios, porque el estadístico Q da un valor de 0,0009, que indica heterogeneidad (Figura 27).

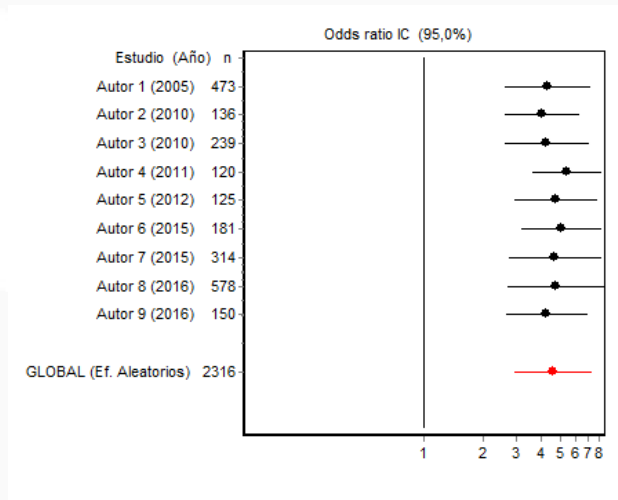
Figura 27
Forest plot para la Odds Ratio



Nota. Elaboración propia.

En la Figura 28 se observa que los intervalos de confianza del 95 % para la odds ratio de todos los estudios se traslapan, por lo que la eliminación de cada uno de ellos no cambia significativamente el resultado global.

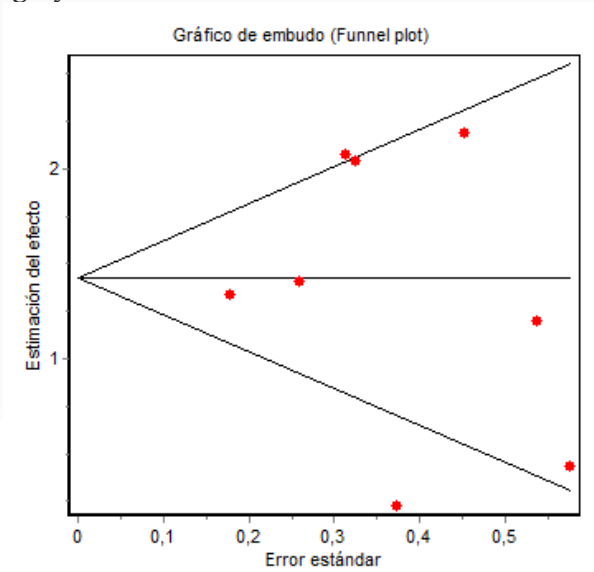
Figura 28
Gráfico de influencias para la Odds Ratio



Nota. Elaboración propia.

El funnel plot, que evalúa sesgo de publicación, pone de manifiesto que los estudios individuales se distribuyen a lado y lado del efecto estimado sin una distribución sesgada evidente. La representación gráfica se corrobora con el estadístico de Begg con un valor p de 0,536, que indica que no hay sesgo de publicación (ver Figura 29).

Figura 29
Funnel plot o gráfico de embudo



Nota. Elaboración propia.

Conclusión: Este metanálisis de estudio de casos y controles

puso al descubierto la relación causal entre el uso previo de antibióticos antituberculosos y la infección por tuberculosis resistente a los fármacos.

Referencias

- Adegbola, C. A., Akhigbe, T. M., Adeogun, A. E., Tvrdá, E., Pizent, A., & Akhigbe, R. E. (2024). A systematic review and meta-analysis of the impact of triclosan exposure on human semen quality [Revisión sistemática y metanálisis del impacto de la exposición al triclosán en la calidad del semen humano]. *Frontiers in Toxicology*, 6, Article 1469340. <https://doi.org/10.3389/ftox.2024.1469340>
- Álvarez-Martínez, H., & Pérez-Campos, E. (2004, julio-agosto). Causalidad en medicina. *Gaceta Médica de México*, 140(4), 467-472. <https://www.scielo.org.mx/pdf/gmm/v140n4/v140n4a18.pdf>
- Aschengrau, A., & Seage G.R. (2020). *Essentials of Epidemiology in Public Health* [Fundamentos de epidemiología en salud pública]. Jones & Bartlett Learning
- Buttignol, M., Bouche, C., Chrétien, M., Taris, N., Ozmen, T., & Mathelin, C. (2025). Hemochromatosis: A Risk Factor for Breast Cancer? Systematic Review and Meta-Analysis [Hemochromatosis: A Risk Factor for Breast Cancer? Systematic Review and Meta-Analysis]. *European Journal of Breast Health*, 21(4), 367-374. <https://doi.org/10.4274/ejbh.galenos.2025.2025-7-1>
- Celentano, D. D., Szklo, M., & Farag, Y. M. K. (2024). Gordis. *Epidemiología* (7 Ed.). Elsevier.
- Cerda, J., Vera, C., & Rada, G. (2013). Odds ratio: aspectos teóricos y prácticos. *Revista Médica de Chile*, 141(10), 1329-1335. <https://dx.doi.org/10.4067/S0034-98872013001000014>
- Dabir, M., Pam, P., Jamali, M., Saba, F., & Ghoreishi, Z. (2025). The association between iron supplementation during pregnancy and the risk of childhood leukemia: a meta-analysis of case-control studies [La asociación entre la suplementación con hierro durante el embarazo y el riesgo de leucemia infantil: un metanálisis de estudios de casos y controles]. *The Journal of Maternal-Fetal & Neonatal Medicine*, 38(1), Article 2474268. <https://doi.org/10.1>

- De Irala, J., Martínez-González, M. A., & Guillén Grima, F. (2001). ¿Qué es una variable de confusión? *Medicina Clínica*, 117(10), 377-385.
- Dekkers, O. M., Vandenbroucke, J. P., Cevallos, M., Renehan, A. G., Altman, D. G., & Egger, M. (2019). COSMOS-E: Guidance on conducting systematic reviews and meta-analyses of observational studies of etiology [COSMOS-E: Orientación sobre la realización de revisiones sistemáticas y metanálisis de estudios observacionales de etiología]. *PLoS Medicine*, 16(2), Article e1002742. <https://doi.org/10.1371/journal.pmed.1002742>
- Ducasse, D., Holden, R. R., Boyer, L., Artéro, S., Calati, R., Guillaume, S., Courtet, P., & Olié, E. (2018). Psychological Pain in Suicidality: A Meta-Analysis [Dolor psicológico en la conducta suicida: un metanálisis]. *The Journal of Clinical Psychiatry*, 79(3), Article 16r10732. <https://doi.org/10.4088/JCP.16r10732>
- Glasziou, P., Irwig, L., Bain, C., & Colditz, G. (2001). *Systematic Reviews in Health Care. A Practical Guide* [Systematic Reviews in Health Care. A Practical Guide]. Cambridge University Press.
- Hernández-Avila, M., Garrido, F., & Salazar-Martínez, E. (2000). Sesgos en estudios epidemiológicos. *Salud Pública de México*, 42(5), 438-446. <https://www.saludpublica.mx/index.php/spm/article/view/6262/7481>
- Judd, N., Hughes, K., Bellis, M. A., Hardcastle, K., & Amos, R. (2023). Is parental unemployment associated with increased risk of adverse childhood experiences? A systematic review and meta-analysis [¿Se asocia el desempleo parental con un mayor riesgo de experiencias adversas en la infancia? Una revisión sistemática y un metanálisis]. *Journal of Public Health*, 45(4), 829-839. <https://doi.org/10.1093/pubmed/fdad069>
- Lazcano-Ponce, E., Salazar-Martínez, E., & Hernández-Avila, M. (2001). Estudios epidemiológicos de casos y controles. Fundamento teórico, variantes y aplicaciones. *Salud Pública de México*, 43(2), 135-150. <https://saludpublica.mx/index.php/spm/article/download/6302/7559>
- Morgan, R. L., Whaley, P., Thayer, K. A., & Schünemann, H. J. (2018). Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes [Identificación del PECO: un

marco para formular buenas preguntas para explorar la asociación entre la exposición ambiental y de otro tipo con los resultados de salud]. *Environment International*, 121(1), 1027-1031. <https://doi.org/10.1016/j.envint.2018.07.015>

Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., Currie, M., Lisy, K., Qureshi, R., Mattis, P., & Mu, P. (2017). Chapter 7: Systematic reviews of etiology and risk [Capítulo 7: Revisiones sistemáticas de etiología y riesgo]. In E. Aromataris, & Z. Munn (Eds.), *Joanna Briggs Institute Reviewer's Manual*. The Joanna Briggs Institute. <https://reviewersmanual.joannabriggs.org/>

Signal, V., Huang, S., Sarfati, D., Stanley, J., McGlynn, K. A., & Gurney, J. K. (2018). Dairy Consumption and Risk of Testicular Cancer: A Systematic Review [Consumo de lácteos y riesgo de cáncer testicular: una revisión sistemática]. *Nutrition and Cancer*, 70(5), 710-736. <https://doi.org/10.1080/01635581.2018.1470655>

Silva, E. P., Lemos, A., Andrade, C., & Ludermit, A. B. (2018). Intimate partner violence during pregnancy and behavioral problems in children and adolescents: a meta-analysis [Violencia de pareja durante el embarazo y problemas de conducta en niños y adolescentes: un metanálisis]. *Jornal de Pediatria*, 94(5), 471-482. <https://doi.org/10.1016/j.jped.2018.01.007>

STROBE. (2025). *What is STROBE?* [¿Qué es STROBE?]. <https://www.strobe-statement.org/>

Tregear, S., Reston, J., Schoelles, K., & Phillips, B. (2009). Obstructive sleep apnea and risk of motor vehicle crash: systematic review and meta-analysis [Apnea obstructiva del sueño y riesgo de accidente automovilístico: revisión sistemática y metanálisis]. *Journal of Clinical Sleep Medicine*, 5(6), 573-581. <https://pubmed.ncbi.nlm.nih.gov/20465027/>

Capítulo 4. Etiología de la enfermedad: síntesis de la literatura de estudios de cohorte

1. Características de los estudios de cohorte

Los estudios de cohorte son investigaciones observacionales analíticas útiles para la identificación de factores de riesgo y factores pronósticos de una enfermedad. En este tipo de estudio epidemiológico, el investigador actúa como un observador (estudio observacional) de eventos relacionados con la exposición “natural” de las personas a sustancias nocivas o saludables. Esta exposición puede darse a través de los hábitos personales (el consumo de alcohol o dietas ricas en grasas), la ocupación (trabajadores expuestos al asbesto), el lugar de residencia (cerca a altas concentraciones de material particulado PM 2,5), entre otras (Aschengrau & Seage, 2020; Celentano et al., 2024). En este sentido, el diseño clásico de los estudios de cohorte parte de la conformación de dos grupos, uno expuesto al potencial factor de riesgo en estudio y otro grupo no expuesto; ambos grupos se siguen a lo largo del tiempo para determinar la incidencia de los resultados de salud (Aschengrau & Seage, 2020; Celentano et al., 2024).

En este capítulo se describe el diseño prospectivo en el que los participantes se agrupan en función de la exposición pasada o actual, y se los sigue para observar los resultados de interés en el futuro. No obstante, también es posible encontrar estudios de cohorte con diseños retrospectivos (en el pasado) o ambidireccionales (un componente retrospectivo y otro prospectivo). El diseño también puede incluir más de dos grupos en función de la exposición, por ejemplo, sin exposición, baja exposición y altos niveles de exposición.

Al comienzo del seguimiento, los participantes de la cohorte están expuestos al potencial factor de riesgo o al factor de riesgo en estudio, pero aún no han desarrollado el evento de interés. Durante el período de seguimiento, los grupos expuestos y no expuestos¹⁸ son monitoreados en tiempo previamente fijados según el horizonte clínico del evento bajo estudio; por ejemplo, cada 3 meses, cada 6

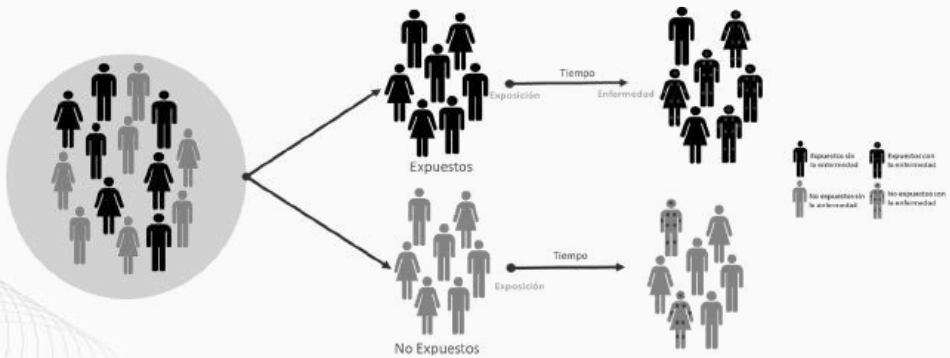
¹⁸ El diseño de este tipo de estudio se puede elaborar para una exposición o para varias.

meses, cada año, etc., para evaluar el desenlace. Al final del período de seguimiento, una proporción de la cohorte expuesta y una proporción del grupo no expuesto (se espera que sea mucho menor) habrán experimentado el evento en estudio. El desenlace para evaluar puede ser la aparición de una enfermedad, las complicaciones, la muerte o incluso efectos positivos en la salud. Así que la duración del seguimiento dependerá del desenlace de interés; por ejemplo, para enfermedades infecciosas el periodo de seguimiento puede ser muy corto, pero para enfermedades como la diabetes, las enfermedades cardiovasculares o el cáncer, el periodo de seguimiento puede requerir de varios años (Aschengrau & Seage, 2020; Celentano et al., 2024) (ver Figura 30).

En este punto es importante tener en cuenta que entre más largo sea el período de seguimiento, más difícil será mantener el contacto con los sujetos del estudio y la investigación será más costosa; y en caso de esperar una alta proporción de pérdidas, se puede considerar el diseño de casos y controles explicado en el capítulo anterior (Aschengrau & Seage, 2020; Celentano et al., 2024).

Figura 30

Representación gráfica de un estudio de cohorte



Nota. Elaboración propia.

La elección del grupo expuesto en un estudio de cohorte depende de la hipótesis a investigar, la frecuencia de la exposición y de aspectos relacionados con la viabilidad (facilidad de seguimiento, disponibilidad de registros médicos, presupuesto). Con base en estas características es posible seleccionar una cohorte especial; es decir, aquella en la que se estudian los efectos de exposiciones poco comunes (ocupacionales a un químico particular, desastres naturales, procedimientos médicos); o una cohorte general para exposiciones

comunes (uso de anticonceptivos orales, tabaquismo, alcoholismo). Un ejemplo de una cohorte especial son los sobrevivientes de la bomba atómica de Hiroshima y Nagasaki en la Segunda Guerra Mundial, en quienes se estudia los efectos de la radiación. Por su parte, un ejemplo de una cohorte general se encuentra en el estudio de Framingham, en el cual los investigadores estudian los factores de riesgo para enfermedades cardiovasculares (Aschengrau & Seage, 2020).

En adición al punto anterior, es importante tener en consideración que **la cohorte puede ser variable o fija dependiendo de la exposición al factor de riesgo. Las cohortes variables se relacionan generalmente con exposiciones que dependen de los estilos de vida como fumar cigarrillos o beber alcohol, en las cuales una persona puede inicialmente tener el hábito de fumar, luego lo abandona por un tiempo y posteriormente retoma el hábito tabáquico. Mientras que las cohortes fijas se definen por la exposición a factores que no cambian con el tiempo como someterse a un procedimiento quirúrgico, tener un hijo o estar presente en un desastre.** Por ejemplo, el caso que vimos anteriormente de los sobrevivientes de la bomba atómica representa de manera adecuada la definición de cohorte fija (Aschengrau & Seage, 2020; Celentano et al., 2024).

La medida epidemiológica típica en esta modalidad de investigación es el riesgo relativo (RR). Para calcularlo, los investigadores tienen en cuenta la incidencia acumulada (o proporción de incidencia) de la enfermedad o el evento en el grupo expuesto ($a/a + b$) y la incidencia acumulada entre los no expuestos ($c/c + d$). La razón entre las incidencias acumuladas en ambos grupos (riesgo en expuestos y riesgo en no expuestos) será el riesgo relativo. La ocurrencia de la enfermedad generalmente se mide usando la incidencia acumulada, las tasas de incidencia y el cálculo del tiempo persona, éste último implica determinar el tiempo de seguimiento de cada persona en el contexto de cada estudio (Celentano et al., 2024).

Antes de exponer la interpretación del riesgo relativo es importante tener presentes las siguientes consideraciones sobre las medidas de incidencia. Tradicionalmente la tasa de incidencia se define como el número de casos nuevos¹⁹ de un evento de interés (enferme-

¹⁹ Es necesario remarcar que nuevo NO corresponde a nuevo caso para el sistema de vigilancia o un nuevo registro de un evento de salud. Es común ver informes de ministerios de salud u otras instituciones que presentan estadísticas de casos incidentes refiriéndose a los nuevos individuos diagnosticados o reportados al sistema de vigilancia; en este caso NO se puede confundir caso incidente con un registro de primera vez. Esto es particularmente en eventos crónicos, donde los casos incidentes deben tener al menos dos mediciones; la primera que demuestre que el individuo era sano o negativo, y una posterior con resultado positivo. En eventos agudos o mediciones de laboratorio que evidencien infecciones o procesos nuevos (por ejemplo, la medición de IgM) no es tan determinante esta aclaración.

dad) que se desarrolla en un periodo de tiempo en la población en riesgo. La tasa de incidencia usualmente se multiplica por 100, por 1000 o cualquier múltiplo de 10. Sin embargo, en sentido estricto cuando no se considera el tiempo de exposición en el denominador, esta definición corresponde a una proporción (el numerador está incluido en el denominador) y no sería una tasa, sino una proporción de incidencia que se conoce como incidencia acumulada. Las tasas son aquellas que en el numerador tienen el número de individuos positivos y en el denominador el tiempo-persona que correspondería a la suma del tiempo de seguimiento de cada individuo estudiado. En este caso, la tasa de incidencia, conocida como densidad de incidencia, sería el cociente de los casos nuevos dividido el tiempo-persona de todos los estudiados (sumatoria del tiempo de seguimiento de los positivos y los negativos).

Lo anterior se ilustra con el siguiente ejemplo: un estudio de cohorte incluyó a 200 personas que fueron seguidas durante seis meses para evaluar la incidencia y los factores de riesgo de un evento. En el primer mes de seguimiento, dos voluntarios resultaron positivos, en el segundo mes se presentaron tres casos, en el tercer mes se registraron seis casos, en el cuarto mes se hallaron siete casos, y en el quinto mes otros ocho voluntarios fueron positivos. Asumiendo que el seguimiento de cada paciente terminó cuando resultó positivo o al finalizar el tiempo del estudio (mes 12), se tendrían las siguientes medidas de incidencia:

1. Incidencia acumulada o proporción de incidencia: $(26/200) \times 100 = 13 \%$. El denominador corresponde a los 200 voluntarios del estudio y el numerador a los 26 casos positivos que se registraron en los diferentes meses del estudio.
2. Tasa o densidad de incidencia: 26/1.138 meses-persona. El numerador corresponde a los 26 casos positivos y el denominador a los tiempos de seguimiento que aportaron los 200 voluntarios así:
 - Los dos primeros positivos aportaron 2 meses (cada uno aporta un mes).
 - Los tres casos siguientes aportaron 6 meses (dos meses cada uno).
 - Los seis casos siguientes aportaron 18 meses (tres meses cada uno).
 - Los siete casos siguientes aportaron 28 meses (cuatro meses cada uno).
 - Los últimos ocho casos aportaron 40 meses (cinco meses cada

uno).

- Los 174 voluntarios que fueron negativos al final del estudio aportaron 1.044 meses de seguimiento (12 meses cada uno).

La tasa de incidencia de $2,28 \times 100$ meses-persona (26/1.138) se puede interpretar así: en una cohorte de 100 persona se presenta un caso cada 2,28 meses o en la población de referencia de la cohorte cada 100 meses se esperan 2,28 personas positivas para el evento. En este tipo de cálculos se asume que el riesgo del evento es constante a lo largo del tiempo, por ello la tasa indica la velocidad de aparición de casos.

Volviendo al riesgo relativo, recordemos la estructura clásica de la Figura 31

Figura 31

Cálculo de riesgos relativos e incidencia

		Enfermedad		
		Enfermo	No enfermo	
Factor de riesgo	Expuesto	a	b	a + b
	No expuesto	c	d	c + d
		a + c	b + d	

Riesgo en expuestos = $\frac{a}{a+b}$	Riesgo relativo = $\frac{\text{riesgo en expuestos}}{\text{riesgo en no expuestos}}$
Riesgo en no expuestos = $\frac{c}{c+d}$	Incidencia = $\frac{a+c}{a+b+c+d}$

Nota. Elaboración propia.

El riesgo relativo puede tomar valores entre 0 e infinito. El 1 indica el valor nulo, es decir, que la enfermedad o el evento ocurre en una proporción de incidencia similar en el grupo expuesto y en el grupo no expuesto. Los valores superiores a 1 indican que el evento se presenta con mayor frecuencia en los expuestos que en los no expuestos, mientras que los valores inferiores a 1 indican que el evento es menor en los expuestos. La definición de factor de riesgo, además de la medida epidemiológica, implica contemplar otros criterios (antecedencia de la exposición, gradiente en la relación dosis respuesta, plausibilidad biológica, fuerza de asociación) (Álvarez-Martínez & Pérez-Campos, 2004); no obstante, debido a que en esta modalidad de investigación se identifican los nuevos casos de enfermedad (in-

cidentes) en la medida en la que van ocurriendo, se establece con mayor claridad o confianza el nexo causal (la antecedenencia) entre la exposición y la enfermedad, por lo que valores mayores a 1 generalmente permiten catalogar la exposición evaluada como de riesgo, y los menos de 1 como factor protector (Celentano et al., 2024) (ver Figura 32).

Figura 32

Representación gráfica de los posibles valores que puede tomar los riesgos relativos



Nota. Elaboración propia.

Con respecto al grupo de comparación, el ideal consiste en tener exactamente los mismos individuos en el grupo de expuestos, pero si no hubieran estado expuestos. No obstante, es imposible que la misma persona esté expuesta y no expuesta simultáneamente, por lo que los investigadores se esfuerzan por seleccionar un buen “contrafactual” (Aschengrau & Seage, 2020; Celentano et al., 2024); es decir, por elegir un grupo de personas lo más similar posible al grupo expuesto, y al hacerlo asegurarse de que la incidencia de la enfermedad en ambos grupos sea idéntica en el caso que la exposición no tenga ningún efecto. Por ejemplo, si los hombres tuvieran un mayor riesgo que las mujeres de desarrollar una enfermedad, es necesario que los grupos expuestos y no expuestos tengan la misma proporción de hombres y mujeres, sumado a la igualdad en potenciales confusores (ya descritos) de la asociación o hipótesis en estudio.

Para decidir cuándo desarrollar un estudio de cohortes es importante tener en cuenta los siguientes criterios. Primero, contar con evidencia empírica de estudios descriptivos en los que se demuestre una asociación (estadística) entre un factor y el desarrollo de una enfermedad o evento en salud. Además, de tener elementos teóricos que respalden una posible relación causal entre la exposición y el

desenlace (Celentano et al., 2024).

Segundo, que la enfermedad o evento en salud sea de alta frecuencia (mayor al 20 %) porque si se realiza un estudio de cohorte para una enfermedad rara, puede ser necesaria una población de estudio muy grande para poder encontrar un número suficiente de individuos que desarrollen la enfermedad (Celentano et al., 2024).

Tercero, que se trate de una enfermedad aguda o de un evento con desenlace rápido, porque recordemos que la población de estudio debe ser seguida durante un período de tiempo para determinar si se desarrolla el resultado de interés. Si se eligen enfermedades crónicas o en las que es necesario una exposición continua y por largos periodos, entonces se requerirá de una fuente de financiación permanente por varios años, además se corre el riesgo que el investigador no pueda sobrevivir hasta el final del estudio (Celentano et al., 2024).

Cuarto, que la hipótesis de estudio permita evaluar diferentes desenlaces ocasionados por una única exposición, porque las cohortes se conforman en relación con la exposición o no a un potencial factor de riesgo y se siguen durante un periodo de tiempo hasta que se vaya presentando el desenlace; no obstante, ese desenlace puede ser único (muerte) o múltiple (diferentes complicaciones de una misma enfermedad). Para finalizar, es necesario tener en cuenta un aspecto logístico no menos importante, y es contar con una cohorte dispuesta a contribuir en el estudio en la que las tasas de abandono o pérdida de seguimiento sean bajas (Celentano et al., 2024). En la Tabla 20 se resumen las características diferenciales de un estudio de cohortes.

Tabla 20
Características distintivas de un estudio de cohorte

Característica	Estudio de Cohorte
Objetivo principal	Analizar la relación entre una exposición y la aparición posterior de una enfermedad o resultado de salud.
Tipo de estudio	Observacional, prospectivo.
Dirección del estudio	Se siguen los participantes de manera prospectiva para observar el desarrollo de la enfermedad o el evento de interés.
Selección de participantes	Los participantes se dividen en dos grupos: expuestos y no expuestos al potencial factor de riesgo.

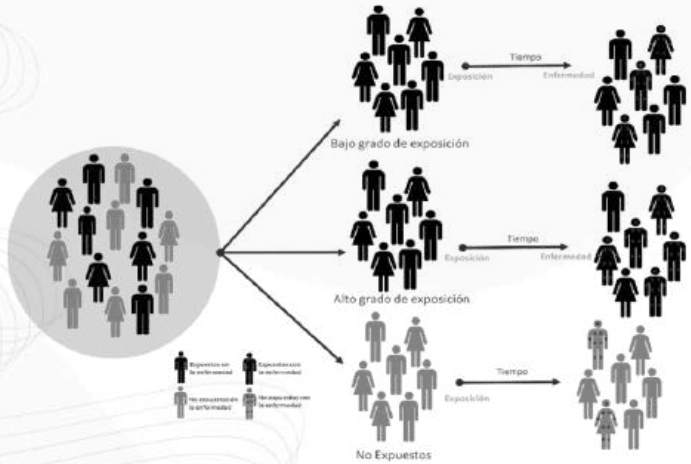
Muestra	La muestra se calcula en función del riesgo relativo esperado.
Medición de la exposición	La exposición se mide antes del inicio del estudio, y se hace seguimiento prospectivo de los participantes durante un período de tiempo.
Costo	Generalmente alto, debido al seguimiento a los participantes y a la necesidad de recopilar datos en varios momentos en el tiempo.
Tiempo requerido	A largo plazo, ya que se requiere seguir a los participantes durante un período significativo para observar el desarrollo de la enfermedad o evento.
Principal medida de asociación	Riesgo relativo (RR).
Ventajas	Permite estudiar relaciones causales; proporciona datos sobre la incidencia de enfermedades.
Desventajas	Costoso y largo en duración; pérdidas de seguimiento pueden introducir sesgos; no siempre es práctico para enfermedades de baja ocurrencia.

Nota. Elaboración propia.

2. Variantes de los estudios de cohortes

Estudio de cohortes múltiples: En esta variante se utilizan varias cohortes que se conforman en función de la magnitud de la exposición a un potencial factor de riesgo, de manera que se puede establecer la relación dosis respuesta (García et al., 2020) (ver Figura 33).

Figura 33
Representación gráfica de los estudios de cohortes múltiples

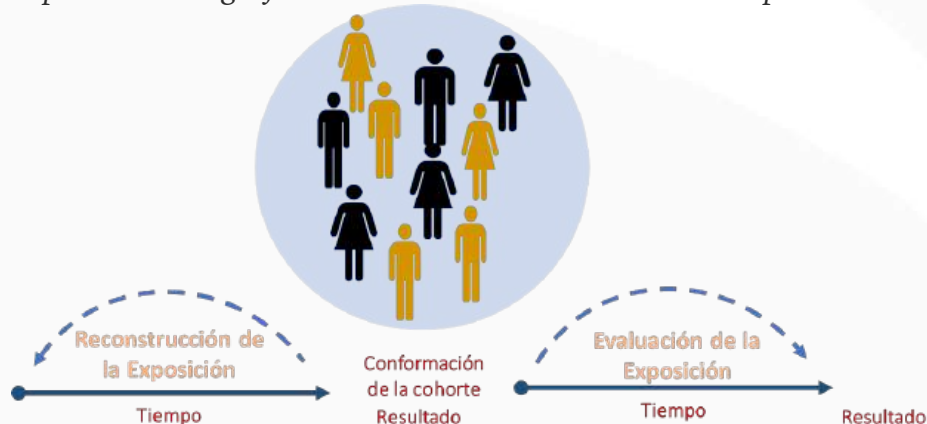


Nota. Elaboración propia.

Estudio de cohortes ambispectivo: En esta modalidad se conforma una cohorte y se recolecta información sobre la exposición a un potencial factor de riesgo desde el momento cero hacia el pasado (retrospectivo) y hacia el futuro (prospectivo) (García et al., 2020) (ver Figura 34).

Figura 34

Representación gráfica de los estudios de cohortes ambispectivos



Nota. Elaboración propia.

Estudios de cohortes dinámicas: en esta modalidad los individuos pueden cambiar la exposición durante el seguimiento; por ejemplo, individuos que tienen el hábito de consumir una cajetilla de cigarrillos diaria pueden pasar a consumir dos cigarrillos a la semana e incluso dejar de consumir cigarrillos (García et al., 2020).

Estudio de cohortes abiertas: en este tipo de estudios se acepta la entrada y salida de nuevos sujetos de la investigación durante el seguimiento, de manera que el número de participantes es variable (García et al., 2020).

Otras variantes de los estudios de cohortes se pueden consultar en el artículo de García et al. (2020).

3. Validez interna y externa en los estudios de cohortes

Al igual que en los capítulos anteriores, la calidad metodológica

es un tópico importante en los estudios de cohortes y por eso es necesario reconocer los aspectos que pueden afectarla. Entre los más relevantes se cuentan: la selección de la cohorte, el seguimiento, la medición de la exposición y el desenlace.

- **Selección de la cohorte**

Los estudios de cohorte deben incluir un número suficiente de individuos para contar con un poder o potencia estadística²⁰ aceptable que permita determinar si una asociación entre dos variables realmente existe o que en caso de no hallar relación se pueda estar seguro de este resultado (que no existe un error mayor a 0,20). Para esto, es necesario determinar el tamaño o número de individuos necesarios a incluir en el grupo expuesto y no expuesto. Este procedimiento se hace con base en el riesgo relativo esperado entre los grupos comparados (utilizando un estudio previo), la confianza (usualmente del 95 %) y la precisión (puede estar entre 1 % y 5 %).

En este punto es importante recordar que el tamaño de los grupos expuestos y no expuestos se calcula para maximizar la potencia estadística y no tienen la intención de imitar la frecuencia de la enfermedad o evento en salud en la población general; esto demuestra la importancia de disponer de buena evidencia en el nivel descriptivo (Aschengrau & Seage, 2020). Ahora bien, para la selección del número mínimo de participantes calculado en el paso anterior, los investigadores deben garantizar que las características de quienes aceptan participar y de quienes rechazan la participación en la investigación no difieran de manera importante, ya que si lo hacen podrían conducir a inferencias equivocadas con respecto a la exposición y el evento.

Por otro lado, los investigadores deben esforzarse en que además de la exposición no haya diferencias importantes entre el grupo expuesto y no expuesto. En las investigaciones experimentales esto se garantiza con la asignación del tratamiento al azar; no obstante, en los estudios de cohorte, al tratarse de investigaciones observacionales, no es posible asignar exposiciones al azar, de manera que las diferencias entre los grupos pueden minimizarse seleccionando cuidadosamente los grupos expuestos y no expuestos. De hecho, **el grupo**

²⁰ La potencia estadística representa la capacidad de una prueba estadística de detectar diferencias cuando estas realmente existen, es el complemento del error β que es indicar que no existen diferencias estadísticas que realmente sí existen. Generalmente se exige que una comparación de grupos tenga una potencia mínima del 80 %. La potencia estadística es influenciada por: a) el tamaño del efecto a detectar (a mayor tamaño del efecto, mayor potencia y necesidad de un menor tamaño de muestra para un nivel de potencia dado), b) el tamaño de la muestra (a mayor tamaño de la muestra, mayor potencia), y c) la variabilidad del evento de interés (a mayor variabilidad, menor potencia o necesidad de un mayor tamaño de muestra para evaluar el efecto).

de comparación ideal consistiría en exactamente los mismos individuos si no hubieran estado expuestos, lo que mencionamos anteriormente como “contrafactual”. Dado que ese grupo ideal es imposible en la práctica, los investigadores recurren a técnicas de emparejamiento individual a partir de características demográficas y clínicas que, además de la exposición, puedan influir el desenlace en estudio (Aschengrau & Seage, 2020).

- **El seguimiento**

El seguimiento de las cohortes se ha vuelto más difícil en los últimos años debido al incremento en la movilidad de la población. Este tema es importante porque las tasas de seguimiento deben ser lo más altas posible para garantizar la validez del estudio. Si bien no existe una tasa de seguimiento que garantice por sí misma la validez, hay consenso entre los investigadores que es suficiente contar con tasas de seguimiento superiores al 90 %; sin embargo, teniendo en cuenta que este criterio podría ser muy exigente, hay autores que recomiendan que se aceptan valores entre el 90 % y el 70 % (Aschengrau & Seage, 2020).

En este sentido, se recomienda que se haga una corrección al tamaño de la muestra, dependiendo de la proporción de pérdidas esperadas, lo cual implicaría agregar entre el 10 % y el 30 % al tamaño de muestra inicial. La pérdida del seguimiento se produce cuando un participante ya no desea continuar en un estudio o no puede ser localizado. Dada su importancia, los investigadores tratan de reducir las pérdidas de seguimiento a través de diferentes estrategias que incluyen la recolección de información exhaustiva referente a varias fuentes de contacto; por ejemplo, la dirección del domicilio actual, número de teléfono fijo, número de celular, dirección de correo electrónico, los nombres, direcciones y números de teléfono de familiares y amigos, etc. En los últimos años se ha popularizado establecer un contacto regular con los participantes por correo electrónico, ya que estas comunicaciones ayudan a los participantes a “vincularse” más con el estudio y a estar en contacto permanente con los investigadores. Sin embargo, el seguimiento es un proceso laborioso y desgastante que implica grandes inversiones de tiempo y dinero; pero es fundamental para garantizar la validez externa del estudio (Aschengrau & Seage, 2020).

Si las pérdidas durante el seguimiento es un fenómeno co-

mún en la investigación, se disminuye el tamaño de la muestra de manera considerable, y con ella la potencia estadística, e incrementa la probabilidad de cometer un error de tipo beta²¹.

Un problema adicional con las pérdidas de seguimiento se encuentra en que los que se pierden pueden diferir de manera importante de los que se rastrean con éxito. Además, si las personas que se pierden tienen más probabilidades de haber desarrollado la enfermedad en estudio sesgaría los resultados. Esto conduce a un sesgo de selección y puede derivar en que las tasas de incidencia en cada grupo sean difíciles de interpretar (Celentano et al., 2024). No obstante, esta limitación podría subsanarse indagando con los participantes si la no continuidad en el estudio se relaciona con el evento o con factores externos como el cambio de país o muerte por causas diferentes a las del estudio.

- **Medición de la exposición y de los desenlaces**

Los investigadores de un estudio de cohortes recurren a varias fuentes para obtener información sobre la medición de las exposiciones y de los desenlaces. Estas fuentes incluyen registros médicos, laborales, encuestas, exámenes físicos, pruebas de laboratorio, monitoreo ambiental, entre otras. Algunas de estas fuentes son preexistentes y otras están diseñadas específicamente para el estudio en ejecución. Las principales fuentes preexistentes son los registros de historias clínicas, porque allí se puede encontrar el historial de exposición de un participante a posibles fuentes de efectos adversos para la salud. Las ventajas de utilizar los registros médicos como fuente de información incluyen un bajo costo y un alto nivel de precisión y detalle con respecto a una enfermedad y su tratamiento; sin embargo, tienen la desventaja de que a menudo falta información sobre variables importantes y la duración de la exposición. En estos casos, es conveniente combinar la información obtenida a través de la historia clínica con encuestas que reúnan la información faltante. Los cuestionarios son particularmente útiles para obtener información sobre las características del estilo de vida que no se registran de manera rutinaria en la historia clínica; sin embargo, es importante tener en cuenta que estos cuestionarios se basan en la capacidad de los participantes para recordar información y, en ocasiones, puede ser difícil para ellos recordar exposiciones que sean de interés para

²¹ Como se indicó en la definición del poder estadístico, un error de tipo Beta se define como la probabilidad de no detectar diferencias estadísticamente significativas en un estudio cuando estas realmente existen. Un error de este tipo podría ocurrir si las pérdidas durante el seguimiento son mayores a la contempladas en la corrección del tamaño de muestra

el estudio (Aschengrau & Seage, 2020).

Por otro lado, hay investigadores que miden la exposición basándose únicamente en datos recopilados específicamente para la investigación. Esto implica someter a los participantes a entrevistas, exámenes físicos, pruebas de laboratorio y otras pruebas, cada cierto tiempo (por ejemplo, cada 6 meses) y durante toda la duración del estudio. En algunos casos, también se recopilan datos de exposición utilizando biomarcadores a través de mediciones biológicas en sangre, orina, huesos y uñas. Los datos recopilados específicamente para un estudio son de mayor calidad que los de los registros existentes, pero suelen ser más costosos (Aschengrau & Seage, 2020).

En todo caso, el investigador debe garantizar que la evaluación de la exposición no sea imperfecta; es decir, que algunas personas expuestas se clasifiquen erróneamente como no expuestas y viceversa. Igualmente, es recomendable seguir procedimientos comparables para determinar los resultados entre los grupos expuestos y no expuestos. Para esto, se puede utilizar una técnica de cegamiento de quienes recolectan la información (similar a los ensayos clínicos) de manera tal que los investigadores desconozcan el grupo expuesto y no expuesto, así como los desenlaces esperados. Este procedimiento garantiza que los esfuerzos de seguimiento sean comparables y la verificación de los desenlaces no estén sesgadas por el investigador (Aschengrau & Seage, 2020; Celentano et al., 2024). Además, los investigadores deben desarrollar definiciones de desenlaces estándar que excluyan el mayor número de interpretaciones posibles, sean específicas y no den lugar a ambigüedades. En el mismo sentido, se debe asegurar que las pruebas diagnósticas utilizadas en ambos grupos sean sensibles (capaces de detectar casos de enfermedad) y específicas (capaces de excluir a personas sin enfermedad). Finalmente, la técnica de cegamiento también se recomienda para las personas que analizarán los datos, así se evita que introduzcan sus prejuicios involuntariamente en sus análisis y en la interpretación de los hallazgos del estudio (Celentano et al., 2024).

A pesar de algunas limitaciones expuestas en los párrafos previos y de la complejidad en términos logísticos y presupuestales que implica desarrollar un estudio de cohorte, esta modalidad de investigación resulta particularmente útil para obtener información sobre los múltiples efectos en salud derivados de una exposición única, cuando la exposición es rara o poco común, y cuando la enfermedad o desenlace es de alta frecuencia. Además, este tipo de estudio es menos vulnerable a sesgos y, por su carácter prospectivo, permite

establecer con mayor claridad el nexo entre la exposición y la enfermedad (Celentano et al., 2024).

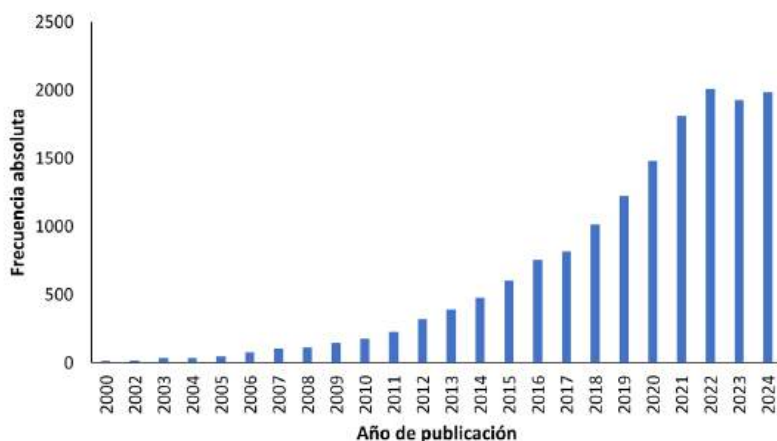
4. Cuándo sistematizar la evidencia de este tipo de estudio. Por qué combinar los resultados

Para llevar a cabo un estudio de cohorte se debe tener alguna idea *a priori* de que una exposición se asocia con una enfermedad, un evento en salud o una complicación y, por tanto, vale la pena investigarla. En consecuencia, una revisión sistemática de estudios de cohorte se debe llevar a cabo cuando hay un acervo teórico y empírico que indica una posible asociación entre una exposición y un desenlace en salud, pero existen divergencias en varios estudios individuales sobre el vínculo entre esa exposición y ese desenlace, sobre la magnitud de la asociación entre el factor de riesgo y la enfermedad, y sobre la dosis respuesta; es decir, sobre el hecho que diferentes niveles de exposición pueden condicionar la gravedad o acortar el intervalo entre la exposición y la enfermedad. Además, es pertinente realizar una revisión sistemática de esta modalidad de investigación cuando los estudios individuales presentan bajos tamaños de muestra por grandes pérdidas durante el seguimiento, ya que al combinar los resultados de varias investigaciones se aumenta la potencia estadística, la validez externa y se reduce la probabilidad de error tipo beta (Celentano et al., 2024).

Teniendo en cuenta estas características, diferentes investigadores han emprendido revisiones sistemáticas de estudios de cohorte. Una búsqueda en PubMed con los términos (meta-analysis[Title]) AND (cohort studies[Title/Abstract]) recupera 12760 resultados, mientras que con la estrategia (systematic review[Title]) AND (cohort studies[Title/Abstract]) se encuentran 15860 publicaciones relacionadas con la síntesis de la literatura de estudios de cohortes entre el 2000 y el 2024 (ver Figura 35).

Figura 35

Frecuencia de publicaciones en PubMed sobre revisiones sistemáticas de estudios de cohortes entre los años 2000 y 2024



Nota. Elaboración propia.

Esas publicaciones incrementan año tras año y se han realizado en temas relacionados con la depresión y la mortalidad por todas las causas (Chan et al., 2025), tratamiento antibiótico y la incidencia de falla renal aguda (Mohammad et al., 2025), el vínculo entre esquizofrenia y cáncer de seno (Mahdavinoor et al., 2024), el uso de cannabis y depresión (Lev-Ran et al., 2014), tabaquismo y nefropatía diabética (Zhu et al., 2025). Además, esta modalidad de investigación también se ha utilizado para identificar factores pronósticos de recurrencia de episodios psicóticos (Fusar-Poli et al., 2016), de la razón neutrófilos-eosinófilos en pacientes con cáncer (Liu et al., 2025), del marcador CXCL12 en pacientes con cáncer (Samarendra et al., 2017), de marcadores moleculares como JAK2, ASXL1 y EZH2 en pacientes con mielofibrosis (Lee, Hsieh et al., 2018), entre otros.

4.1 Características del protocolo de búsqueda y selección de estudios en la síntesis de estudios de cohortes

4.1.1 La estrategia de búsqueda

La estrategia de búsqueda de una revisión sistemática de estudios de cohorte con el objetivo de identificar factores etiológicos debe contener los elementos de la pregunta PEO (Population, Exposure, Outcome) (Moola et al., 2017). Es decir, se debe especificar la población, la exposición y el desenlace o evento en salud. De manera que, si al investigador le interesa sistematizar la evidencia disponible sobre, por ejemplo, el riesgo de parto prematuro (outcome) en

mujeres (population) fumadoras (exposure), la estrategia de búsqueda deberá contener esas palabras clave, así como los términos homólogos. Esta estrategia de búsqueda sería suficiente si el investigador desea realizar una revisión sistemática desde el enfoque amplio y la información disponible sobre el tópico de interés es exigua; no obstante, si al realizar una búsqueda exploratoria se encuentra que la información sobre el tema es prolífica e inmanejable para el equipo de investigación, se recomienda realizar una búsqueda desde el enfoque restrictivo, agregando palabras clave relacionadas con el tipo de estudio objeto de la síntesis como “cohorte” o “prospectivo” (ver Tabla 21)²².

Tabla 21
Frecuencia absoluta de los artículos identificados en cada base de datos y con una estrategia de búsqueda de enfoque amplio

PubMed Title/Abstract			
AND	Cigarrillo	Tabaco	Tabaquismo
Parto premature	n	n	n
Parto pretérmino	n	n	n
Nacimiento pretérmino	n	n	n
Scielo Title/Abstract			
AND	Cigarrillo	Tabaco	Tabaquismo
Parto premature	N	n	n
Parto pretérmino	N	n	n
Nacimiento pretérmino	N	n	n
Google Scholar Title			
AND	Cigarrillo	Tabaco	Tabaquismo
Parto premature	n	n	n
Parto pretérmino	n	n	n
Nacimiento pretérmino	n	n	n

Nota. Elaboración propia.

En el caso que el objetivo sea sintetizar la evidencia sobre estu-

²² En este punto es importante aclarar que esta recomendación ayuda a los estudiantes a optimizar esta parte de la revisión (la fase inicial de identificación de la guía PRISMA). Sin embargo, se debe recordar que el enfoque o alcance se refiere a la pregunta que orientó la revisión sistemática; por lo que se espera que en la escritura del problema de investigación el estudiante haya analizado la factibilidad de su revisión, en la cual se debería tener un número estimado de los estudios disponibles o un panorama general de la evidencia previa (en estudios originales e incluso otras revisiones), de tal forma que no deba cambiar el enfoque-alcance de la pregunta de la revisión durante su ejecución.

Si alguien llega a la estrategia de búsqueda, y debe cambiar al enfoque restrictivo, más que un cambio en la estrategia de búsqueda debería ajustar la problematización. En el caso de hallar un número muy elevado de textos, en lugar de cambiar a un enfoque-alcance restrictivo (lo cual ameritaría cambiar problema y pregunta de investigación-revisión) se debería buscar opciones para reducir este inmanejable número agregando términos o usando los filtros que le den bases de datos.

dios relacionados con factores pronósticos, algunos autores sugieren utilizar la estrategia PICOTS (Population, Intervantion, Comparator, Outcome, Timming, Setting) (Riley et al., 2019). En este contexto, la “P” de la población y la “O” del resultado siguen siendo en gran medida las mismas que para estudios etiológicos, pero la “T” se refiere al factor pronóstico en estudio y la “C” a otros factores pronósticos que pueden considerarse como comparadores. La “T” denota el tiempo en que se deben medir o evaluar el factor de interés y el periodo de tiempo en que ese factor predice el resultado o desenlace. Finalmente, La “S” se refiere al contexto en el que se utilizará el factor pronóstico; por ejemplo, en contextos de atención primaria, secundaria o en hospitales de alto nivel de complejidad (Riley et al., 2019).

Algunos autores sugieren el diseño de una única estrategia de búsqueda utilizando la combinación de todos los términos con operadores booleanos como AND, OR, NOT; sin embargo, entre más extensa sea la estrategia, mayor será la probabilidad de errores por inclusión u omisión de paréntesis o de operadores booleanos en lugares incorrectos, lo que afecta la reproducibilidad. En este sentido, se recomienda utilizar búsquedas simples y cortas.

4.1.2 Los criterios de inclusión

En la definición de los criterios de inclusión se debe verificar que los estudios recuperados con la estrategia de búsqueda respondan a los elementos de la pregunta PEO. Así, pueden especificarse tres o cuatro criterios de inclusión. En el primero se verifica que la población corresponda a los individuos de interés (adultos, jóvenes, ancianos, hombres, mujeres, etc). En el segundo se confirma que la exposición concierna a la que los investigadores quieren estudiar (Tabaquismo, alcohol, material particulado PM2,5, asbesto, etc). En el tercero los investigadores deben cerciorarse de que el desenlace (incidencia de enfermedad, muerte, complicaciones) sea igual al que orienta la investigación; no obstante, como los estudios de cohortes pueden tener diferentes desenlaces, bastaría con que incluya el principal.

Finalmente, dependiendo de la estrategia de búsqueda elegida se pueden recuperar artículos como revisiones narrativas, otras revisiones sistemáticas, estudios descriptivos o, incluso, ensayos clínicos, de manera que es recomendable agregar un criterio de inclusión que indique que se incluirán estudios originales (para eliminar revisio-

nes narrativas, cartas al editor) y un criterio que indique que solo interesan los estudios de cohorte (para eliminar otra tipología de estudios originales).

4.1.3 Los criterios de exclusión

Los criterios de exclusión se aplican a los artículos que cumplieron con los criterios descritos en la sección anterior y su finalidad es eliminar publicaciones que puedan conducir a sesgos en los resultados o comprometer la validez externa. Con base en lo anterior, estos criterios se formulan teniendo en cuenta la selección de la cohorte, el seguimiento, la medición de la exposición y los desenlaces. En la selección de la cohorte se debe procurar excluir estudios en los que los individuos elegidos para el seguimiento difieran sustancialmente de aquellos quienes no se incluyeron o que el contrafactual no sea comparable con la cohorte expuesta. En relación con el seguimiento se recomienda excluir estudios con pérdidas mayores al 70 %, sin que hayan contemplado la corrección de las pérdidas en el cálculo del tamaño de la muestra, ya que usualmente quienes no logran seguirse con éxito son quienes tienen peores desenlaces en salud²³. Con respecto a la medición de la exposición y los desenlaces, se sugiere excluir investigaciones en las que la medición de ambas variables se haya realizado con el uso de técnicas con bajo desempeño diagnóstico. Por último, se deberán excluir investigaciones que no reporten la medida de asociación epidemiológica (riesgo relativo) para los factores estudiados, ni los datos que permitan calcularla.

4.1.4 La extracción de la información

Para la extracción de la información se recomienda diseñar un archivo en Microsoft Excel. La información para extraer en cada artículo es similar a lo visto en los capítulos anteriores, de manera que ese archivo en Excel debe contener variables relacionadas con el estudio, con la cohorte y con el análisis de la información. Las variables relacionadas con el estudio incluyen: los autores, el año de publicación y el lugar de realización (ciudad, país). Las variables

²³ En este punto se debe hacer un análisis más detallado de la explicación de pérdidas, dado que las revisiones sistemáticas subsanan los problemas de poder estadístico, principalmente los generados por bajos tamaños de muestra. No obstante lo anterior, este criterio de exclusión no remite a un problema de poder estadístico, sino de sesgos en los casos que no se pudo conocer o explicar la causa del abandono del estudio (y se tienen elementos para sospechar que dicho abandono se relaciona con presentar el evento de estudio).

También es importante aclarar que en las revisiones sistemáticas que no usan la calidad metodológica como criterio de elegibilidad, sino como variables que se presentarán en los resultados, no sería necesario aplicar este criterio de exclusión, sino que esto se informaría en los resultados y se tendría presente en los análisis de sensibilidad del metanálisis o en la realización de meta-regresiones.

relacionadas con los participantes deben contener características demográficas y clínicas de la cohorte expuesta y no expuesta, tales como la edad, el sexo, las comorbilidades, el tamaño de la muestra en cada grupo. En la exposición deberá especificarse el tipo de exposición, la intensidad y la duración. Por su parte, la medición del desenlace estudiado debe ir acompañada de la definición de ese desenlace, la gravedad y la técnica diagnóstica utilizada para cuantificarlo.

Para extraer los datos relacionados con el análisis de la información, se sugiere utilizar cuatro columnas diferentes del archivo de Excel en el que se diligencie la información de la tabla de 2 x 2; es decir, el número de personas expuestas que desarrollaron la enfermedad (a), el número de expuestos que no desarrollaron la enfermedad (b), los no expuestos que desarrollaron la enfermedad (c) y los no expuestos que no desarrollaron la enfermedad. En los casos en los que esta información esté disponible, será el insumo para realizar el metanálisis de riesgos relativos. Finalmente, en caso de disponer del RR ajustado es necesario que se consideren en la extracción de la información las variables y métodos de dicho ajuste.

4.2 Evaluación de la calidad metodológica y riesgos de sesgos

El riesgo de sesgos y la evaluación de la calidad metodológica se realiza a los artículos objeto de la síntesis. Para esto existen múltiples guías que pueden ser utilizadas, entre ellas se destacan la guía STROBE (*STrengthening the Reporting of OBservational studies in Epidemiology*) para estudios de cohortes; la guía propuesta por el instituto Jonana Briggs específica de este tipo de estudios (Moola et al., 2017) y la guía de lectura crítica de estudios de cohorte de la red CASPe (Cabello, 2005).

La guía CASPe contiene 11 preguntas y están distribuidas en tres secciones, las preguntas, en su mayoría (excepto las de la sección B), se califican en términos de “Sí cumple”, “No cumple” o “No sé”. Las dos primeras preguntas de la sección A son de eliminación y solo si la respuesta a ellas es afirmativa vale la pena continuar con las demás. Las tres preguntas restantes de esta misma sección dan cuenta de la validez de los resultados. En la sección B se incluyen dos preguntas con la que se describen los resultados del estudio. Finalmente, en la sección C hay cuatro preguntas sobre la aplicabilidad de los resultados que en algunos casos no se tiene en cuenta en la calidad

metodológica (validez interna y externa), sino en la valoración de los usos o la utilidad de la evidencia sistematizada (Tabla 22). La aplicación de esta guía debe ser realizada por dos investigadores de manera independiente y las discrepancias deberán ser resueltas por consenso o, cuando no sea posible, por referencia a un tercer investigador.

Tabla 22
Criterios para evaluar la calidad metodológica de los estudios de cohorte

Criterios para evaluar la calidad metodológica de estudios de cohortes				
Sección	Pregunta	Sí Cumple	No cumple	No sé
Sección A	¿El estudio se centra en un tema claramente definido?			
	¿La cohorte se reclutó de la manera más adecuada?			
	¿El resultado se midió de forma precisa con el fin de minimizar posibles sesgos?			
	¿Han tenido en cuenta los autores el potencial efecto de los factores de confusión en el diseño y/o análisis del estudio?			
	¿El seguimiento de los sujetos fue lo suficientemente largo y completo?			
Sección B	¿Cuáles son los resultados de este estudio?			
	¿Cuál es la precisión de los resultados?			
	¿Le parecen creíbles los resultados?			
Sección C	¿Los resultados de este estudio coinciden con otra evidencia disponible?			
	¿Se pueden aplicar los resultados en su medio?			
	¿Va a cambiar este estudio su decisión clínica?			

Nota. Elaboración propia.

4.3 Análisis de la información

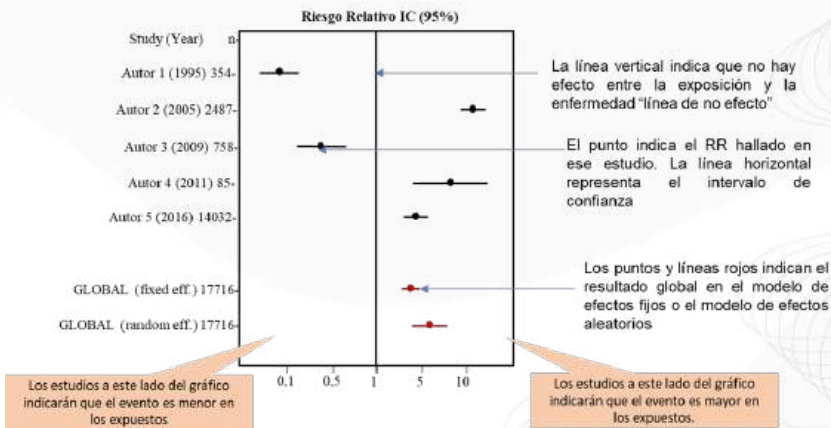
El análisis de la información de una revisión sistemática de estu-

dios de cohortes debe contar con dos secciones: una síntesis cualitativa y una síntesis cuantitativa. En la síntesis cualitativa se presenta la distribución geográfica de las publicaciones, para la cual suele usarse un mapa señalando los lugares en los que se han realizado las investigaciones sobre el tema de interés y aquellos en los que la investigación es escasa o nula. Además, se incluyen una o dos tablas con las características de la cohorte, la exposición y el desenlace. Para describir los resultados de estas tablas, basta con el cálculo de frecuencias absolutas, relativas con sus intervalos de confianza del 95 % y medidas de resumen (posición, dispersión y tendencia central), en los casos en los que sea posible.

En algunas situaciones, cuando hay heterogeneidad en las mediciones (conceptual, no estadística), solo es posible realizar un metanálisis de medidas indirectas comparando algunos riesgos relativos con sus intervalos de confianza del 95 %, pero sin estimar medidas combinadas por divergencias en el diseño o la medición. Cuando tal divergencia conceptual no existe, se realiza el metanálisis “tradicional” que incluiría la evaluación de la heterogeneidad, el sesgo de publicación, el forest plot y el análisis de la sensibilidad.

En términos generales, el forest plot de un metanálisis de estudios de cohortes resume el tamaño del efecto entre la exposición y el desenlace, a partir de riesgos relativos. La línea vertical sobre el 1 representa la ausencia de relación entre la exposición y el evento, los valores superiores a 1 indican que el evento es mayor en el grupo expuesto, mientras que los valores inferiores a 1 indican que el evento es mayor en los no expuestos (ver Figura 36).

Figura 36
Aspecto de un forest plot para un metanálisis de estudios de cohortes.



Nota. Elaboración propia.

Al igual que se comentó en el capítulo de casos y controles, en el metanálisis de estudios de cohortes no se captan los ajustes que se hacen en los análisis multivariados en los estudios individuales, de manera que se debe ser cuidadoso en la interpretación, ya que los hallazgos podrían estar influenciados por potenciales factores de confusión²⁴. Otro aspecto que se debe tener en cuenta es que además del forest plot con la medida global, los investigadores pueden hacer diferentes forest plot en el mismo metanálisis, agrupando las cohortes según características demográficas, clínicas, del gradiente de exposición y de los diferentes desenlaces, los cuales permitirían realizar meta-regresiones.

Referencias

- Álvarez-Martínez, H., & Pérez-Campos, E. (2004, julio-agosto). Causalidad en medicina. *Gaceta Médica de México*, 140(4), 467-472. <https://www.scielo.org.mx/pdf/gmm/v140n4/v140n4a18.pdf>
- Aschengrau, A. & Seage G.R. (2020). *Essentials of Epidemiology in Public Health* [Fundamentos de epidemiología en salud pública]. Jones & Bartlett Learning.
- Cabello, J. B. (2005). Plantilla para ayudarte a entender Estudios de Cohortes. En CASPe, *Guías CASPe de Lectura Crítica de la Literatura Médica* (pp.23-27). CASPe.
- Celentano, D. D., Szklo, M., & Farag, Y. M. K. (2024). *Gordis. Epidemiología* (7 Ed.). Elsevier.
- Chan, J. K. N., Solmi, M., Lo, H. K. Y., Chan, M. W. Y., Choo, L. L. T., Lai, E. T. H., Wong, C. S. M., Correll, C. U., & Chang, W. C. (2025). All-cause and cause-specific mortality in people with depression: a large-scale systematic review and meta-analysis of relative risk and aggravating or attenuating factors, including antidepressant treatment [Mortalidad por todas las causas y por causas específicas en personas con depresión: una revisión sistemática a gran escala y un metanálisis del riesgo relativo y los factores agravantes o atenuantes, incluido el tratamiento antidepresivo]. *World Psychiatry Association*, 24(3), 404-421. <https://>

²⁴ Recordar hacer una extracción de este tipo de información para evaluar la posibilidad de hacer meta-regresiones por estas variables.

- Fusar-Poli, P., Cappucciati, M., Bonoldi, I., Hui, L. M., Rutigliano, G., Stahl, D. R., Borgwardt, S., Politi, P., Mishara, A. L., Lawrie, S. M., Carpenter, W. T., & McGuire, P. K. (2016). Prognosis of Brief Psychotic Episodes: A Meta-analysis [Pronóstico de episodios psicóticos breves: un metanálisis]. *JAMA Psychiatry*, 73(3), 211-220. <https://doi.org/10.1001/jamapsychiatry.2015.2313>
- García, N., Manterola, C., Guerrero, E., & Navarrete, O. (2020). Estudios de cohortes especiales. *Revista de Cirugía*, 72(2), 171-178. <https://dx.doi.org/10.35687/s2452-45492020002690>
- Lee, Y. C., Hsieh, C. C., Lee, Y. L., & Li, C. Y. (2018). Molecular Markers and Prognosis of Myelofibrosis in the Genomic Era: A Meta-analysis [Marcadores moleculares y pronóstico de la mielofibrosis en la era genómica: un metanálisis]. *Clinical Lymphoma, Myeloma & Leukemia*, 18(9), 558-568. <https://doi.org/10.1016/j.clml.2018.06.004>
- Leite, F., Nascimento, G. G., Scheutz, F., & López, R. (2018). Effect of Smoking on Periodontitis: A Systematic Review and Meta-regression [Efecto del tabaquismo en la periodontitis: una revisión sistemática y metarregresión]. *American Journal of Preventive Medicine*, 54(6), 831-841. <https://doi.org/10.1016/j.amepre.2018.02.014>
- Lev-Ran, S., Roerecke, M., Le Foll, B., George, T. P., McKenzie, K., & Rehm, J. (2014). The association between cannabis use and depression: a systematic review and meta-analysis of longitudinal studies [La asociación entre el consumo de cannabis y la depresión: una revisión sistemática y metanálisis de estudios longitudinales]. *Psychological Medicine*, 44(4), 797-810. <https://doi.org/10.1017/S0033291713001438>
- Liu, T., Gao, X., Geng, C., Yao, Y., & Zhang, G. (2025). Prognostic role of the baseline neutrophil eosinophil ratio in cancer patients: a meta-analysis and systematic review [Prognostic role of the baseline neutrophil eosinophil ratio in cancer patients: a meta-analysis and systematic review]. *World Journal of Surgical Oncology*, 23(1), Article 334. <https://doi.org/10.1186/s12957-025-03981-1>
- Mahdavinoor, S. M. M., Kargar-Soleimanabad, S., Bordbari, A. H., Mollaei, A., Seddigh, L., Sarihi, S., FiroozKhojastefar, R., Dehbozorgi, R., & Farhang, S. (2024). Investigating the Relationship

between Schizophrenia and Incidence Risk of Breast Cancer: A Systematic Review and Meta-Analysis of Cohort Studies [Investigación de la relación entre la esquizofrenia y el riesgo de incidencia de cáncer de mama: una revisión sistemática y un metanálisis de estudios de cohortes]. *Iranian Journal of Psychiatry*, 19(4), 441-452. <https://doi.org/10.18502/ijps.v19i4.16559>

Mohammad, S., Ghazal, H., Rahimeh, W., Khan, M., Al Balas, M., & El-Dahiyat, F. (2025). Comparative Risk of Acute Kidney Injury with Piperacillin-Tazobactam Plus Teicoplanin Versus Piperacillin-Tazobactam Plus Vancomycin: A Systematic Review and Meta-Analysis [Riesgo comparativo de lesión renal aguda con piperacilina-tazobactam más teicoplanina versus piperacilina-tazobactam más vancomicina: una revisión sistemática y un metanálisis]. *Drug Safety*. <https://doi.org/10.1007/s40264-025-01611-z>

Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., Currie, M., Lisy, K., Qureshi, R., Mattis, P., & Mu, P. (2017). Chapter 7: Systematic reviews of etiology and risk [Capítulo 7: Revisiones sistemáticas de etiología y riesgo]. In E. Aromataris, & Z. Munn (Eds.), *Joanna Briggs Institute Reviewer's Manual*. The Joanna Briggs Institute. <https://reviewersmanual.joannabriggs.org/>

Riley, R. D., Moons, K., Snell, K., Ensor, J., Hooft, L., Altman, D. G., Hayden, J., Collins, G. S., & Debray, T. (2019). A guide to systematic review and meta-analysis of prognostic factor studies [Una guía para la revisión sistemática y el metanálisis de estudios de factores pronósticos]. *BMJ*, 364, Article k4597. <https://doi.org/10.1136/bmj.k4597>

Samarendra, H., Jones, K., Petrinic, T., Silva, M. A., Reddy, S., Soonawalla, Z., & Gordon-Weeks, A. (2017). A meta-analysis of CXCL12 expression for cancer prognosis [Un metanálisis de la expresión de CXCL12 para el pronóstico del cancer]. *British Journal of Cancer*, 117(1), 124-135. <https://doi.org/10.1038/bjc.2017.134>

Zhu, H., Li, L., Liu, S., & Li, J. (2025). Smoking and diabetic nephropathy: An updated systematic review and meta-analysis [Tabaquismo y nefropatía diabética: una revisión sistemática y un metanálisis actualizados]. *Journal of Diabetes Investigation*, 16(3), 442-450. <https://doi.org/10.1111/jdi.14385>

Capítulo 5. El tratamiento: síntesis de la literatura de estudios de intervención

1. Características de los estudios de intervención

En los capítulos anteriores hemos visto diferentes tipos de estudio que buscan describir la frecuencia de una enfermedad, identificar los factores de riesgo y la forma de diagnosticarla; sin embargo, tanto en el ámbito clínico como en el campo de la salud pública, es fundamental identificar alternativas para modificar la historia natural de la enfermedad, bien sea a través de la prevención o del tratamiento.

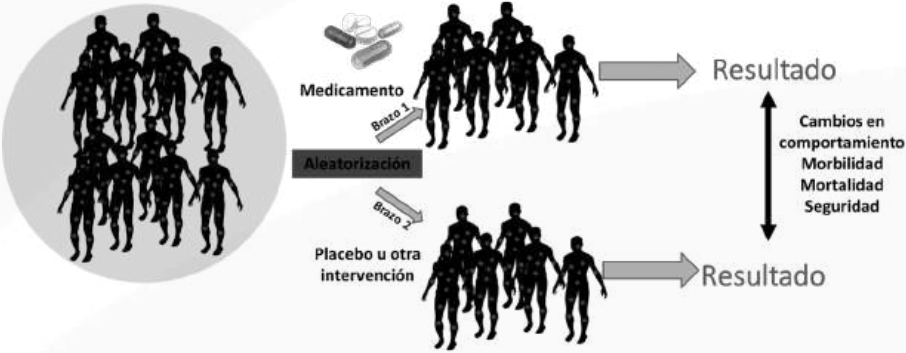
Seleccionar una alternativa preventiva o terapéutica implica el diseño de estudios que permitan cuantificar la eficacia y los efectos secundarios de las intervenciones que se llevarán a cabo, de modo que se pueda elegir la mejor opción disponible. En este sentido, los estudios de intervención son un tipo de investigación que se conduce en condiciones controladas por el investigador, con el propósito de evaluar el efecto que genera una variable sobre los sujetos de estudio (que pueden ser modelos celulares, animales o personas). En este diseño, el investigador asigna activamente la exposición a los participantes, es decir, que manipula la variable independiente de mayor relevancia. **El hecho que el investigador controle o manipule una variable hace que el diseño deje de nombrarse como observacional y se clasifique como experimental (Celentano et al., 2024), lo que convierte a los estudios de intervención en el diseño metodológico por excelencia para establecer relaciones de causalidad.**

En la medicina moderna, existe un consenso más o menos generalizado en que el ideal de los estudios de intervención son los ensayos clínicos controlados aleatorizados. El diseño clásico de este tipo de estudios consiste en seleccionar a un grupo de individuos de enfermos, dividirlos en dos grupos homogéneos, administrar el tratamiento a probar a un grupo y dar a los del otro grupo un producto inocuo o placebo (o el tratamiento convencional). La mayoría de los medicamentos que conocemos hoy en día para el tratamiento de las enfermedades crónicas no transmisibles, así como para las

enfermedades infecciosas, se han probado con diseños de este tipo (Celentano et al., 2024) (ver Figura 37).

Figura 37

Representación grafico del diseño de un estudio de intervención



Nota. Elaboración propia.

No obstante, antes de evaluar la eficacia de los medicamentos en el grupo de enfermos, los investigadores deben realizar estudios previos de intervención en modelos celulares y modelos animales; estos estudios se conocen como experimentales pre-clínicos²⁵. Al obtener resultados satisfactorios en las etapas pre-clínicas, se realizan estudios en humanos que se clasifican en ensayos clínicos fase 1, fase 2, fase 3 y fase 4 (Umscheid et al., 2011).

Los ensayos clínicos fase 1 tienen el objetivo de evaluar la seguridad, la farmacocinética, la farmacodinámica y la interacción de diferentes medicamentos. Usualmente, el diseño de este tipo de estudios consiste en seleccionar a un pequeño grupo de participantes, que pueden ser sanos o enfermos, a quienes se le aplicará la intervención de interés. Luego, se les sigue de manera cercana para buscar cualquier evidencia de toxicidad del medicamento. Los ensayos fase 1 también son conocidos como estudios de dosis escalonadas o de farmacología humana (Umscheid et al., 2011).

Los ensayos fase 2 tienen el objetivo de evaluar la eficacia preliminar del medicamento, pero también otorgan información clave para planear los ensayos clínicos fase 3 con datos de la seguridad, la determinación de la dosis óptima del medicamento, la frecuencia de

²⁵ En este contexto se usa el término preclínico para agrupar una amplia diversidad de estudios cuya unidad de análisis no es un paciente o individuo, sino un proceso fisiológico. Los más común es disponer de evidencia previa procedente de modelos preclínicos en animales, pero dependiendo del mecanismo de acción, la efectividad y seguridad esperada, los riesgos y normativas que regulan diferentes tipos de intervención, se pueden usar otros modelos in vivo (diferentes a los modelos animales) como los estudios en células o tejidos; modelos ex vivo, organoides y microchips, estudios in sillico o modelos in vitro.

la dosis y la vía de administración. Este tipo de estudio se conduce en un grupo pequeño de voluntarios con la enfermedad y la eficacia preliminar se establece al comparar con datos de estudios previos o comparando frente a un grupo control. No obstante, el hecho de que este tipo de estudios se realice en un grupo pequeño de participantes, impide que se establezca la eficacia real; por esa razón se conocen como estudios exploratorios terapéuticos (Umscheid et al., 2011).

En sentido estricto, los ensayos fase dos se focalizan en evaluar-determinar la dosis-respuesta y ofrecer datos sobre la seguridad de la intervención a corto plazo. Como valor agregado, brindan datos iniciales de la respuesta terapéutica como un concepto diferente al de eficacia, dado que la respuesta terapéutica inicial solo es un indicio de la acción de la intervención, sin que implique una mejoría clínica real o significativa. Los ensayos fase dos también pueden agrupar desenlaces favorables en marcadores biológicos, observaciones clínicas iniciales y otras evidencias de un posible efecto clínico. La eficacia implica la demostración de un beneficio clínico significativo en condiciones controladas (a diferencia de la efectividad que corresponde a condiciones no controladas o de la práctica clínica habitual).

En los ensayos clínicos fase 3 el objetivo es evaluar la eficacia del medicamento y estimar la incidencia de los efectos adversos más comunes. Los ensayos fase 3 se realizan en un grupo grande de individuos que representan las personas con la enfermedad a quienes se les aplicaría el tratamiento en la práctica clínica real (Umscheid et al. 2011). Una de las características distintivas de este tipo de estudios es el balance en los grupos que recibirán los tratamientos, procedimiento que se lleva a cabo a través de la aleatorización, de manera que los individuos que participan en este tipo de ensayos tienen la misma probabilidad de pertenecer al grupo que recibirá el tratamiento nuevo que se está probando o al grupo que recibirá el tratamiento estándar o convencional (en algunos casos se da un placebo, pero por cuestionamientos éticos, la asignación de placebos es cada vez menos común²⁶). **Con la aleatorización se eliminan diferencias sistemáticas o variables confusoras entre las personas que recibirán ambos tratamientos.** Dependiendo de la expectativa que se tenga del tratamiento en evaluación, el ensayo clínico fase 3

²⁶ El uso de placebos cada vez es menos frecuente, pero se puede argumentar su uso en los ensayos que garanticen que durante su ejecución este grupo control no tendría avance del estado de su enfermedad o cuando no existen otros tratamientos. Por ejemplo, en algunas enfermedades infecciosas de curso lento el uso de placebo se podría justificar en términos éticos e incluso este grupo permitiría evaluar la proporción de curación espontánea (autocuración por los mecanismos inmunes del paciente).

puede clasificarse como un estudio de superioridad (el tratamiento nuevo será mejor que el tratamiento estándar) o de no inferioridad (el tratamiento nuevo no será inferior al tratamiento estándar).

Otra característica de los ensayos fase 3 es el cegamiento, lo que significa que ni los pacientes (simple ciego) ni los investigadores (doble ciego) conocen si están recibiendo el tratamiento nuevo o el tratamiento convencional. Por último, los ensayos fase 3 deben hacer explícito el tipo de análisis que llevarán a cabo, los cuales se clasifican en análisis por intención de tratar y análisis por protocolo. **En el análisis por intención de tratar se tiene en cuenta el número de individuos que se asignó a cada brazo al inicio del estudio, de manera que se respeta el resultado de la aleatorización enunciado previamente. Por su parte, en el análisis por protocolo los sujetos son evaluados teniendo en cuenta el número de individuos que recibió cada tratamiento al final de estudio**²⁷. El análisis por protocolo introduce sesgos al estudio, porque al tener en cuenta solo los que terminan el ensayo se pierden los beneficios de la aleatorización inicial; del mismo modo, puede llegar a sobreestimar la seguridad del ensayo porque los individuos que se retiran pueden ser justo aquellos que presentaron mayor incidencia de efectos adversos. Los ensayos fase 3 también se conocen como estudios de confirmación terapéutica, eficacia comparativa o ensayo clínico pivotal (Umscheid et al., 2011).

Los ensayos clínicos fase 4 se realizan después de la comercialización de los medicamentos con el objetivo de evaluar la aparición de efectos adversos menos comunes y la efectividad del tratamiento. Esta fase no se realiza en condiciones controladas, no hay grupo de control y los efectos adversos se reportan directamente por los médicos tratantes o los pacientes, por lo que podrían estar sesgados. A pesar de la importancia de esta fase de los ensayos, algunas estimaciones sugieren que menos de la mitad de los estudios que requieren pasar por esta fase se llevan a cabo. Este tipo de ensayos se conocen como estudios de uso terapéutico, de farmacovigilancia o estudios post comercialización (post-marketing) (Umscheid et al., 2011).

Por último, es importante tener en cuenta que los desenlaces que se evalúan en los ensayos clínicos pueden ser variables clínicas duras, como la disminución en la mortalidad, hospitalización o eventos

²⁷ Por intención de tratar se calcula la eficacia sobre el total de individuos al inicio del estudio; esto implica que el número de sujetos que abandona el estudio (pérdidas de seguimiento) se incluyen como fallas terapéuticas. El numerador sería el número de individuos en quienes la intervención fue eficaz al finalizar el estudio y el denominar todos los sujetos calculados al inicio del estudio (los que finalizaron el estudio y los abandonos). En el cálculo de la eficacia por protocolo, que excluye del cálculo a las pérdidas de seguimiento, el numerador es el mismo del cálculo por intención de tratar, pero el denominador solo tiene en cuenta a los sujetos que terminaron el estudio (resta las pérdidas de seguimiento).

cardiovasculares; desenlaces intermedios, como la alteración (disminución o incremento) de parámetros bioquímicos o fisiológicos (perfil lipídico, glicemia o la presión arterial); y desenlaces centrados en los pacientes, entre los que se incluye la calidad de vida relacionada con la salud o la percepción del dolor.

La medida de asociación típica en los ensayos clínicos controlados es el riesgo relativo (RR), que compara la incidencia del desenlace en el grupo de intervención frente al grupo control. No obstante, en desenlaces cuantitativos se utiliza la diferencia de medias. Otras medidas que se usan con cierta frecuencia son la reducción absoluta del riesgo (RAR) y el número necesario a tratar (NNT), y se calcula como el inverso de la medida anterior ($1/\text{RAR}$); esta medida proporciona una estimación más práctica del efecto de la intervención. Estas medidas, al igual que en otros diseños, deben complementarse con intervalos de confianza y pruebas de significación estadística. Las características principales de este tipo de estudios se resumen en la Tabla 23.

Tabla 23
Características distintivas de un estudio de intervención tipo ensayo clínico

Característica	Ensayo clínico
Objetivo principal	Evaluar la eficacia y seguridad de una intervención.
Tipo de estudio	Experimental.
Dirección del estudio	Prospectivo, se inicia con la intervención y se hace un seguimiento a las pacientes para evaluar el desenlace.
Selección de participantes	Los participantes son seleccionados de acuerdo con la presencia de la enfermedad que se desea intervenir y se realiza una asignación aleatoria al grupo del tratamiento a evaluar y al grupo del tratamiento estándar.
Medición de la exposición	La exposición es asignada por el investigador y puede ser un medicamento, una intervención educativa o una vacuna.
Costo	Costosos, requieren de un equipo grande de investigadores y alto presupuesto para hacer el seguimiento a los participantes.
Tiempo requerido	Largo, ya que después de asignar la intervención, se siguen a los pacientes hasta observar el desenlace.
Principal medida de asociación	Riesgo relativo (RR), Número necesario a tratar (NNT), reducción absoluta del riesgo (RAR) y diferencia de medias.

Ventajas	Alta validez interna, son el diseño ideal para establecer relaciones causales.
Desventajas	Costoso, tiempo prolongo, plantean grandes cuestiones éticas.

Nota. Elaboración propia.

2. Variantes en los estudios de intervención

Los ensayos clínicos controlados aleatorizados han sido objeto de diversas adaptaciones metodológicas que responden a las características particulares de las enfermedades, las intervenciones y los contextos en que se aplican. Una de estas variantes es el ensayo cruzado, en el cual los mismos participantes reciben, en momentos distintos y en orden aleatorizado, tanto la intervención como el comparador. Entre ambos periodos se establece un intervalo de lavado, que permite reducir el efecto residual de la intervención anterior. Este diseño es especialmente útil cuando la condición clínica es estable y el efecto del tratamiento es transitorio. Su principal fortaleza es que cada participante actúa como su propio control, lo que mejora la precisión del estudio al disminuir la variabilidad interindividual. Sin embargo, no es recomendable cuando los efectos de la intervención pueden prolongarse o modificar el curso natural de la enfermedad (Lazcano-Ponce et al., 2004).

Otra variante de este tipo de estudios es el diseño por conglomerados. En este caso, la aleatorización se realiza sobre grupos de personas (como escuelas, hospitales o comunidades) y no sobre individuos. Este diseño es particularmente apropiado cuando se quiere intervenir sobre contextos colectivos o cuando existe alto riesgo de contaminación entre los participantes (Lazcano-Ponce et al., 2004).

Por último, los ensayos pragmáticos buscan evaluar la efectividad de las intervenciones en las condiciones reales de la práctica clínica. A diferencia de los ensayos clínicos aleatorizados fase 3, que enfatizan el control metodológico y la homogeneidad de la muestra, los ensayos pragmáticos incluyen una población más heterogénea y simulan las condiciones habituales de atención en salud.

3. Validez interna y externa en los estudios de intervención

Hemos visto que los ensayos clínicos buscan modificar la historia natural de una enfermedad; por tanto, resulta de especial interés que los resultados de este tipo de investigaciones puedan aplicarse, no solo a quienes hicieron parte del estudio, sino a otras personas que no han participado, lo que se conoce como validez externa. Sin embargo, para que los resultados de un estudio de intervención puedan aplicarse, es necesario garantizar que no se producen errores en la selección de los participantes, la asignación de las intervenciones, la comparabilidad de los grupos, las mediciones, la digitación o el análisis de la información; esto se conoce como validez interna. En la mayoría de las investigaciones se privilegia la validez interna por encima de la externa, debido a que si los resultados no reflejan de manera adecuada lo que ocurre con los individuos estudiados, pues menos aún lo hará con los que no participaron de la investigación.

En el contexto específico de los ensayos clínicos, algunos autores sugieren que en la actualidad existen múltiples recursos para diseñar estudios impecables desde la perspectiva de la validez interna, pero se ha descuidado la validez externa, por lo que los resultados de esas intervenciones resultan irrelevantes para la práctica clínica. La razón principal de esa crítica es que se hacen en condiciones experimentales tan controladas, que hace imposible que los resultados puedan generalizarse a otra población (Gérvás & Pérez Andrés, 2008). En este contexto, es necesario que los estudios de intervención busquen un balance entre la validez externa y la interna. En la literatura se han descrito múltiples sesgos que pueden afectar la validez interna y externa de los ensayos clínicos; sin embargo, los más relevantes comprenden cinco dominios que incluyen: los sesgos de selección, sesgos de realización, sesgo de detección, sesgo de desgaste y sesgo de reporte (Higgins & Green, 2011).

El sesgo de selección se refiere a la presencia de diferencias sistemáticas entre las características sociodemográficas y clínicas de base de los grupos que recibirán los tratamientos y serán comparados, lo que significa que la distribución de factores pronósticos puede ser diferente en ambos grupos. Por ejemplo, al comparar dos intervenciones para pacientes con infección por SARS-Cov-2, el investigador debe asegurarse de que quienes recibirán la intervención 1 no difieren de quienes recibirán la intervención 2, en variables como la edad o la presencia de comorbilidades como diabetes, obesidad, inmunocompromiso, entre otras, ya que si el grupo que recibe el tratamiento que se está probando es más joven y más saludable que el grupo que recibe el tratamiento estándar, entonces la eficacia y la seguridad

del nuevo medicamento será sobrestimada.

El sesgo de realización ocurre cuando hay diferencias sistemáticas en el cuidado que reciben los pacientes con ambos tratamientos y ocurre cuando no hubo un adecuado cegamiento. El hecho que un individuo o el investigador conozca el tipo de tratamiento que está recibiendo puede generar que se extremen los cuidados durante la participación en el estudio generando que el resultado final no sea producto del tratamiento, sino de los demás cuidados que recibió el participante. Pese a lo anterior, hay situaciones clínicas en las que no es posible realizar cegamiento; por ejemplo, en tratamientos relacionados con intervenciones quirúrgicas.

El sesgo de detección también está relacionado con un inadecuado proceso de cegamiento o también ocurre cuando hay diferencias sistemáticas entre los grupos al momento de evaluar el resultado de interés. Para evitarlo debe procurarse que el médico, microbiólogo, bacteriólogo o quien evalúe el resultado de la intervención no conozca el tratamiento que recibió cada uno de los pacientes.

El sesgo de desgaste ocurre cuando hay diferencias sistemáticas entre los grupos debido a las pérdidas de pacientes durante la investigación, lo que puede ocasionar que se pierda el efecto de la aleatorización, disminuya la potencia estadística y los grupos estén incompletos al momento del análisis. Por ejemplo, si un grupo de los individuos que participa en un ensayo clínico tiene peor pronóstico, entonces es posible que la enfermedad les impida asistir a futuras visitas de los investigadores. Del mismo modo, si entre los tratamientos que se están comparando hay uno menos potente que el otro y los pacientes conocen esa información (inadecuado cegamiento), entonces quienes reciben el medicamento menos potente tienen más probabilidades de no asistir a futuras visitas.

Por último, se encuentra el sesgo de reporte que se presenta cuando los resultados de un ensayo clínico no se reportan debido a que los investigadores no encontraron un efecto significativo del tratamiento evaluado. Este tipo de sesgos es relevante en el contexto de las revisiones sistemáticas y los metanálisis, porque al evaluar el total de los estudios publicados se hace evidente que las investigaciones con resultados contrarios a las hipótesis de los investigadores no se están publicando (Higgins & Green, 2011).

4. Indicaciones para sistematizar la evidencia en los estudios de intervención y elegir cuándo combinar los resultados

Los ensayos clínicos, por su diseño experimental, constituyen la fuente de evidencia de mayor jerarquía para responder preguntas sobre la eficacia de una intervención. Sin embargo, los resultados de los estudios pueden diferir en la conclusión sobre la eficacia de una misma intervención debido a variaciones en el tipo de población, la intervención aplicada, los desenlaces medidos o la calidad metodológica. Por ello, las revisiones sistemáticas y metanálisis de ensayos clínicos son indispensables para sintetizar la evidencia disponible, resolver controversias en torno a un tratamiento y guiar la toma de decisiones clínicas y en salud pública.

La sistematización de este tipo de estudios es especialmente relevante cuando: i) existen estudios individuales con resultados contradictorios, ii) se desea obtener una estimación combinada más precisa del efecto que genera la intervención evaluada, iii) se necesita evaluar la consistencia de los hallazgos en distintos contextos o países, y iv) se buscan subgrupos de pacientes que se benefician más de la intervención o que puedan estar en mayor riesgo de desarrollar efectos adversos. Se aumenta la potencia estadística al estimar los efectos globales.

En PubMed, una búsqueda con los términos (meta-analysis[Title]) AND (clinical trial[Title/Abstract]) arroja más de 3.000 publicaciones de esta modalidad de investigación entre los años 2000 y 2024, abarcando terapias farmacológicas, intervenciones quirúrgicas, estrategias educativas y tecnológicas, entre muchas otras (Figura 38). Entre los metanálisis de ensayos clínicos se encuentran estudios sobre el suplemento con vitamina B 12 para el tratamiento profiláctico de la migraña (Chen et al., 2022), tratamientos antibióticos en pacientes pediátricos con apendicetomías no complicadas (Huang et al., 2017), el uso de probióticos en pacientes con síndrome de intestino irritable (Didari et al., 2015), el tratamiento con acupuntura para pacientes con depresión (Chen et al., 2023) y el uso de terapias endoscópicas para pacientes con obesidad (Weitzner et al., 2023), entre otras. Estos trabajos ilustran el amplio espectro de aplicaciones de la síntesis de la evidencia de ensayos clínicos.

Figura 38

Frecuencia de publicaciones en PubMed sobre revisiones sistemáticas de estudios de ensayos clínicos entre los años 2000 y 2024



Nota. Elaboración propia.

4.1 Características del protocolo de búsqueda y selección de estudios en la síntesis de ensayos clínicos

Para revisar sistemáticamente la evidencia de los estudios de intervención, se recomienda seguir la estrategia PICO: Población, Intervención, Comparador y Desenlace. La “P” se refiere a la población con la enfermedad de interés; la “I” hace referencia a la intervención que se desea probar, el tratamiento que será objeto de la evaluación; la “C” será el tratamiento estándar, frente al que se harán las comparaciones. Como se mencionó antes, en el comparador solían utilizarse placebos o productos inocuos; sin embargo, estos son cada vez menos frecuentes por limitaciones de tipo ético. Por último, la “O” incluirá las medidas de eficacia y seguridad.

4.1.1 La estrategia de búsqueda

La estrategia de búsqueda es el pilar de toda revisión sistemática. Debe diseñarse de manera que sea reproducible, sensible y específica. En el caso de los ensayos clínicos, la estrategia debería estar guiada por la pregunta de investigación, siguiendo el formato PICO descrito en la sección anterior. Para construir una buena estrategia de búsqueda de este tipo de estudios, se recomienda iniciar con una búsqueda exploratoria, usando términos clave relacionados con la

condición clínica, la intervención, el comparador y los desenlaces. Esta búsqueda permite identificar sinónimos, términos MeSH, operadores booleanos y campos específicos de la base de datos a los que se restringirá la búsqueda (como [Title/Abstract]).

Posterior a lo enunciado, se sugiere conducir una búsqueda estructurada y exhaustiva en bases de datos como PubMed, Cochrane, Scopus, Web of Science y algunas bases de datos regionales como Scielo. En esta estrategia, se recomienda utilizar una combinación de términos de lenguaje controlado (disponibles en Mesh o Decs) y libres, y aplicar operadores booleanos como OR para palabras homólogas o AND para enlazar los componentes de la pregunta PICO (ver Tabla 24). Algunos estudios también incluyen filtros propios de las bases de datos como “Randomized Controlled Trial [Publication Type]”; sin embargo, su sensibilidad y especificidad es variable, por lo que se desaconseja esta práctica. Adicional a lo anterior, es deseable revisar referencias bibliográficas de los artículos incluidos, a fin de captar artículos que no fueron identificados con la estrategia de búsqueda.

Por último, es importante precisar que en algunos escenarios se sugiere incluir todos los componentes de la estrategia PICO en la búsqueda, con el propósito de obtener resultados muy específicos; no obstante, se recomienda que en fenómenos en los que la evidencia es exigua, no está tan estandarizada la práctica de cuidado o se desean describir los resultados de la intervención frente a distintos comparadores, es suficiente con que la estrategia de búsqueda incluya la enfermedad, la intervención de interés, el desenlace a evaluar y un concepto propio del tipo de estudio como eficacia o seguridad.

Tabla 24
Estrategia de búsqueda básica para una revisión sistemática de estudios de intervención

Ejemplo de una estrategia de búsqueda básica						Base de datos		
Enfermedad de interés		Intervención		Desenlace	Concepto del tipo de estudio	Pubmed	Scielo	Google scholar
Fibromyalgia	AND	Physical activity	AND	Quality of life	Efficacy	n	n	n
		Motor activity				n	n	n
		Exercise				n	n	n

Nota. Elaboración propia.

Una vez se obtienen todos los resultados de la búsqueda en cada base de datos, se envían a una fuente común con el propósito de identificar duplicados y aplicar criterios de inclusión y exclusión. Para realizar estas etapas existen múltiples herramientas, incluso pueden realizar en Excel o un gestor de referencias como Zotero o Mendeley, pero también se han desarrollado herramientas específicas para conducir revisiones sistemáticas, entre las que se destacan Covidence, Rayyan y CADIMA, las cuales se caracterizan porque son útiles y fáciles de usar, incluso para investigadores novatos (Van der Mierden et al., 2019).

4.1.2 Los criterios de inclusión

Al aplicar los criterios de inclusión es importante tener en cuenta si los artículos identificados permiten responder la pregunta que se planteó el investigador. En este sentido, es necesario enunciar aspectos propios de la población como el diagnóstico, el estadio o severidad de la enfermedad, la presencia de comorbilidades, la edad, el sexo, entre otros aspectos de interés. En el mismo sentido, es necesario considerar en los criterios de inclusión las características propias de la intervención, sobre todo su descripción detallada en los artículos a incluir. En algunos casos se sugiere que la intervención sea igual en términos de frecuencia, intensidad, duración, vía de administración; no obstante, también es posible incluir las distintas variantes de la intervención y hacer análisis de subgrupos con las variantes. El tipo de comparador puede considerarse o no al momento de aplicar criterios de inclusión, ya que si se desea evaluar la eficacia del tratamiento de interés frente a todos los comparadores posibles, entonces no sería necesario hacerlos explícitos²⁸. Por el contrario, si el objetivo está dirigido a comparar dos intervenciones específicas, entonces el comparador debería incluirse como parte de los criterios de inclusión. Un criterio de inclusión adicional que puede ser útil en este tipo de estudios es que se mencione específicamente el desenlace a evaluar, esto es mortalidad, mejoría clínica, efectos adversos o calidad de vida. Por último, se deben enunciar que se incluirán exclusivamente ensayos clínicos o el tipo de estudio de intervención que se haya propuesto sistematizar.

4.1.3 Los criterios de exclusión

²⁸ Esta es la opción más usada, dado que las diferencias en el comparador, o incluso su ausencia, se pueden analizar mediante metanálisis de comparaciones indirectas, descrito en el capítulo de los estudios descriptivos.

Los criterios de exclusión se formulan para descartar estudios con alto riesgo de sesgo o con información incompleta para responder la pregunta de investigación. En este sentido, los criterios de exclusión deben dar cuenta de los siguientes aspectos: i) estudios con evidentes problemas metodológicos, como ausencia de aleatorización, seguimientos inadecuados o pérdidas de seguimiento superiores al 30 %; ii) publicaciones secundarias o duplicadas con la misma población; iii) ensayos que no reportan los datos necesarios para el metanálisis, como las medidas de efecto, los riesgo relativo o las medias junto con la medida de dispersión (en el caso de desenlaces cuantitativos); y iv) se debe mencionar la exclusión de artículos a los que no fue posible acceder al texto completo.

En esta fase es importante tratar de equilibrar el rigor metodológico con la amplitud de la evidencia disponible. De manera que, en lugar de excluir estudios con limitaciones metodológicas, puede optarse por evaluar su calidad e incluirlos con análisis por subgrupos o meta regresiones para explorar su influencia en los resultados de los ensayos clínicos. En este caso, solo se aplicarían criterios de exclusión relacionados con una información incompleta o que no es posible acceder al texto completo a pesar de buscar en todas las bases de datos e intentar contactar a los autores. Incluso, es más riguroso no aplicar de manera tan estricta criterios de elegibilidad relacionados con la calidad metodológica, porque el componente siguiente de este tipo de revisión implica evaluar el riesgo de sesgos en este tipo de estudios. Si fuese una recomendación universal excluir estudios por temas de calidad, no tendría sentido explicitar la importancia de aplicar herramientas de evaluación de la calidad metodológica o el riesgo de sesgos en las fases subsiguientes de la revisiones sistemática (ver capítulo siguiente).

4.1.4 La extracción de la información

La extracción de datos de cada artículo debe ser realizada por dos revisores de manera independiente, usando un formulario estructurado para ese fin (por ejemplo, en Excel). Los desacuerdos en la extracción de la información deben resolverse por consenso o mediante la consulta de un tercer revisor. El formulario de extracción debe contener datos generales del estudio como autor, año de publicación, país de realización. Características de la población como el tamaño de la muestra, edad, sexo, enfermedad de interés, criterios diagnósticos, comorbilidades. Características de la intervención y el

comparador, incluyendo el tipo de intervención, dosis, duración del tratamiento, adherencia. Los desenlaces evaluados y la medida utilizada, el tiempo de seguimiento y las técnicas de medición. Finalmente, los datos para los análisis estadísticos como riesgos relativos, medias, desviación estándar, entre otros. Idealmente, debe registrarse el tipo de análisis reportado (intención de tratar vs. por protocolo) y si los resultados fueron ajustados en modelos multivariados.

4.2 Evaluación de la calidad metodológica y riesgos de sesgos

Existen múltiples herramientas para evaluar la calidad metodológica y el riesgo de sesgos de los estudios de intervención; sin embargo, existe un consenso generalizado en que la herramienta más recomendada para evaluar el riesgo de sesgo en ensayos clínicos es RoB 2 (*Risk of Bias* por sus siglas en inglés), desarrollada por la Colaboración Cochrane (Cochrane methods bias, 2025). Esta herramienta permite juzgar la calidad metodológica de cada estudio incluido en una revisión, de acuerdo con los siguientes dominios:

Sesgo derivado del proceso de aleatorización: se refiere a si la asignación de los participantes a los grupos se realizó mediante métodos realmente aleatorios y si no hay diferencias sustanciales entre los grupos que recibirán cada uno de los tratamientos.

Sesgo por no ocultamiento de la asignación: analiza si la secuencia aleatoria estuvo protegida y fue impredecible para los investigadores.

Sesgo por desviaciones de las intervenciones previstas: considera si los participantes y el personal siguieron la intervención asignada sin violaciones importantes, y si el análisis respetó la aleatorización original.

Sesgo por datos incompletos de desenlaces: examina pérdidas durante el seguimiento, abandonos y cómo fueron manejados.

Sesgo en la medición de los desenlaces: evalúa si los evaluadores estaban cegados a la intervención y si los métodos de medición fueron válidos y consistentes.

Sesgo por reporte selectivo de los resultados: analiza si todos los desenlaces preespecificados fueron reportados, o si se omitieron o modificaron según los hallazgos (Sterne et al., 2019).

Cada uno de estos dominios se evalúa en términos de si hay: i)

Bajo riesgo de sesgo, ii) Alguna preocupación con respecto al riesgo de sesgo, o iii) Alto riesgo de sesgo. Finalmente se otorga un juicio global para cada estudio.

Para facilitar la comprensión de los hallazgos, se recomienda presentar los resultados del RoB 2 en tablas y gráficos tipo semáforo o de barras apiladas, que permiten visualizar rápidamente el nivel de riesgo de sesgo en cada dominio por estudio (ver Figura 39). Además, es importante tener presente que esta evaluación no solo cumple una función descriptiva dentro de los resultados de la revisión, sino que puede utilizarse posteriormente para realizar análisis de sensibilidad (excluyendo estudios de alto riesgo) y evaluar el impacto del sesgo mediante meta regresiones

Figura 39
Calificación del riesgo de sesgo de los estudios individuales de acuerdo con la guía RoB 2



Nota. Elaboración propia.

4.3 Análisis de la información

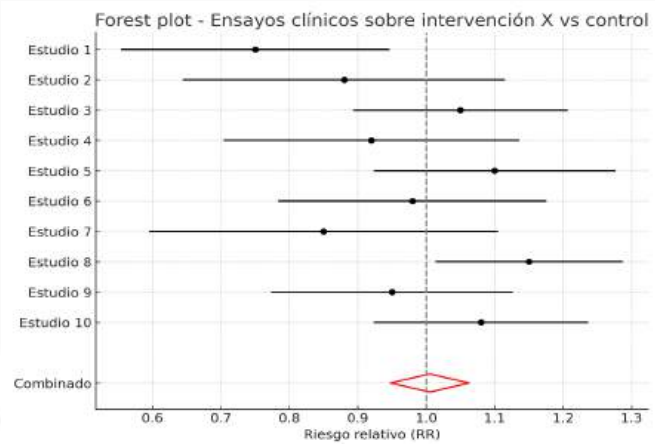
El proceso de análisis debe iniciar con una síntesis clara del proceso metodológico, desde la búsqueda hasta la inclusión final de estudios, especificando cuántos artículos se identificaron, cuántos fueron excluidos en cada etapa y cuáles fueron los motivos de exclusión. Esta información debe representarse gráficamente mediante un diagrama de flujo conforme a los lineamientos PRISMA. Poste-

riormente, se debe elaborar una descripción detallada de las características de los estudios que serán objeto de la síntesis, incluyendo datos sobre las poblaciones intervenidas, los tipos de intervención y comparador, los desenlaces evaluados y la duración del seguimiento en cada estudio. Se recomienda acompañar esta descripción con tablas que presenten frecuencias absolutas y relativas para las variables extraídas.

Sumado a lo anterior, es importante recordar que una revisión sistemática con metanálisis no solo busca sintetizar la evidencia, sino evaluar su calidad. Por esa razón, debe incluirse un resumen narrativo de la calidad metodológica de los ensayos analizados, incorporando los resultados de la evaluación del riesgo de sesgo. En caso de haberse utilizado la herramienta de Cochrane, deben presentar gráficamente los resultados de cada dominio y un resumen general.

La síntesis cuantitativa se lleva a cabo cuando los estudios son suficientemente homogéneos en diseño, población, intervención y desenlaces. En este caso, se hace el metanálisis, donde se estima el efecto combinado para la medida de interés (RR, OR, diferencia de medias, etc.). Los resultados se presentan mediante gráficos tipo forest plot, los cuales muestran los efectos individuales, los intervalos de confianza y la estimación global del efecto (Figura 40).

Figura 40
Forest plot para un riesgo relativo.



Nota. Elaboración propia.

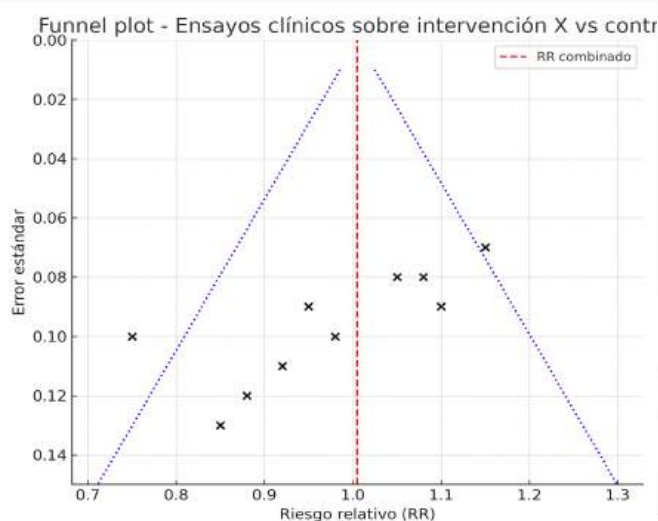
El forest plot muestra que la mayoría de los estudios individuales presentan riesgos relativos con intervalos de confianza que cruzan la unidad, lo que indica que ninguno de ellos demuestra un efecto

estadísticamente mayor o significativo de la intervención frente al control de forma individual. Además, la medida combinada, representada por un rombo rojo, también incluye la unidad en su intervalo de confianza, lo que significa que el efecto global no es mayor. En conjunto, estos resultados sugieren que la eficacia de la intervención fue estadísticamente igual a la del comparador (si el estudio era no superioridad si hay evidencia concluyente de la eficacia), por lo que la elección de la intervención o el comparador debería centrarse en los resultados de seguridad. Si el estudio es de superioridad, se concluye que eficacia entre ambos fue similar, no hay evidencia de la superioridad, pero la elección de la intervención o el comparador debería tener en cuenta los análisis de seguridad y aplicabilidad de cada tratamiento.

La heterogeneidad entre estudios debe evaluarse con el estadístico I^2 , y si se detecta pueden aplicarse análisis de subgrupos o meta regresiones para explorar sus causas. Los hallazgos más relevantes deben exponerse de manera narrativa en el texto, incluyendo una interpretación clínica del efecto combinado. Además, es importante reportar si se identificó sesgo de publicación, por ejemplo, mediante gráficos de embudo (funnel plot; ver Figura 41) y pruebas estadísticas como la de Egger. Todo el análisis debe enfocarse no solo en la significación estadística, sino también en la relevancia clínica de los resultados combinados.

Figura 41

Representación gráfica del análisis de sesgo de publicación en un ensayo clínico



Nota. Elaboración propia.

El funnel plot muestra la dispersión de los efectos individuales (riesgos relativos) en función de su error estándar. La mayoría de los puntos se concentran alrededor de la medida combinada (línea roja discontinua), y su distribución es simétrica dentro de los límites del embudo (líneas azules punteadas), lo que sugiere una baja probabilidad de sesgo de publicación. No se observa una gran asimetría.

5. Ejemplo práctico

Problema de investigación: La fibromialgia es una enfermedad crónica e incapacitante que se caracteriza por dolor generalizado. Para manejar el dolor y mejorar la calidad de estos pacientes, se han explorado múltiples tratamientos farmacológicos y no farmacológicos. Entre los tratamientos no farmacológicos, se destacan las intervenciones con sesiones de ejercicio físico; sin embargo, los resultados de los ensayos clínicos no son concluyentes, pues hay estudios que indican que se mejora considerablemente la calidad de vida de los pacientes, pero otros indican que no hay diferencias con el tratamiento estándar.

Objetivo: Evaluar la eficacia de la actividad física en el mejoramiento de la calidad de vida de pacientes con fibromialgia en la literatura científica mundial.

Método: Se utilizó la estrategia de búsqueda: (((fibromyalgia[Title/Abstract]) AND (exercise[Title/Abstract])) AND (quality of life[Title/Abstract])) AND (efficacy[Title/Abstract]) reemplazando en cada búsqueda la palabra exercise con motor activity, physical activity. La búsqueda se hizo en las bases de datos ScienceDirect, PubMed, Scopus y Scielo con la aplicación de criterios de inclusión y exclusión. La calidad de vida se evaluó con el instrumento FIQ que genera un puntaje de 0 a 100; a mayor puntaje, mejor calidad de vida. Teniendo en cuenta que la medida de efecto o desenlace es cuantitativa, se hizo un metanálisis de diferencia de medias.

Resultados: Se identificaron 7 artículos que evaluaron la calidad de vida de pacientes con fibromialgia, antes y después de la intervención con ejercicio físico, comparados con un grupo de control que recibió solo el tratamiento estándar para la enfermedad. A los artículos se les extrajo la información de la media y la desviación estándar del puntaje del FIQ (calidad de vida) en los dos momentos evaluados, tal como se muestra en la Tabla 25.

Tabla 25
Extracción de la información en un estudio experimental clínico

Autor	Grupo	Tamaño de muestra	Puntaje FIQ	
			Pretrata- miento	Postratamiento
Autor 1	Grupo Experimental	13	29,8±14,1	51,8±19,5
	Grupo control	14	33±13,5	37,2±17,7
Autor 2	Grupo Experimental	43	36,5±14,4	48,8±17,3
	Grupo control	31	37,2±12,6	36,3±12,8
Autor 3	Grupo Experimental	36	34,2±15,4	55,9±19,4
	Grupo control	32	35,2±18,2	50,9±19,4
Autor 4	Grupo Experimental	18	39±21	45±17
	Grupo control	18	43±17	42±18
Autor 6	Grupo Experimental	20	34,9±11,5	49,8±17,2
	Grupo control	21	32,5±14	32,8±13,9
Autor 7	Grupo Experimental	15	53,1±13,1	58,2±13,3
	Grupo control	16	45,1±12,6	52,2±15,9
Autor 8	Grupo Experimental	34	23,5±12,1	49,5±16,6
	Grupo control	32	26,3±11,3	23±10

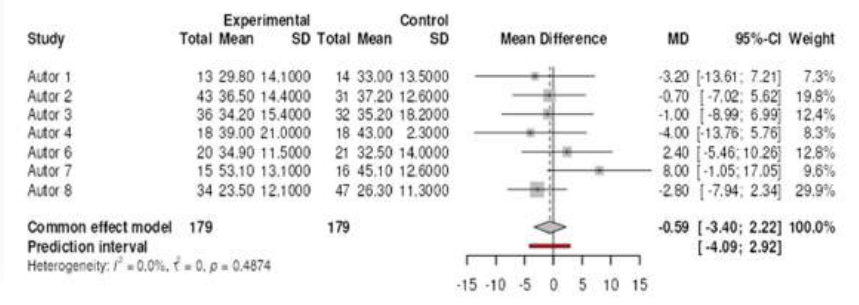
Nota. Elaboración propia.

La información de la Tabla 25 se introdujo en la herramienta de simulación de metanálisis disponible online (<https://www.bioestadistica.uma.es/app/metaanalisis/metaanalisis-num/>) y se realizó el metanálisis para comparar la calidad de vida entre los dos grupos evaluados, antes de iniciar con la intervención del ejercicio físico. El propósito de este análisis es demostrarle al lector que no hay diferencias significativas en la calidad de vida de los grupos a comparar, al inicio del estudio, que puedan condicionar el resultado final.

En el metanálisis se observa que no hay heterogeneidad (I^2 fue de 0,0) por lo que se realiza un análisis bajo el modelo de efectos fijos. Los resultados ponen de manifiesto que en ninguno de los estudios individuales hubo diferencias significativas en la calidad de vida de los dos grupos antes de iniciar la intervención del ejercicio físico, debido a que todos los intervalos de confianza atraviesan el 0 (constituye la línea de no diferencia en un estudio de diferencia de medias). Coherente con eso, la medida combinada muestra que la deferencia de medias fue de -0,59 con un intervalo de confianza (IC 95 % -7,94 a 2,34) que también atraviesa el 0, ratificando la ausencia de diferencia entre los grupos (ver Figura 42).

Figura 42

Forest plot para un ensayo clínico de diferencia de medias. (Antes de la intervención)

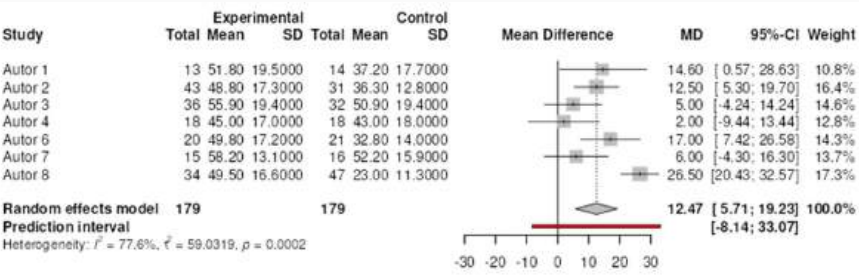


Nota. Elaboración propia.

Una vez se aseguró la homogeneidad de los grupos, se realizó un metanálisis del puntaje de calidad de vida después de la intervención con el ejercicio físico. Los resultados muestran alta heterogeneidad (I^2 fue de 77,6 %), por lo que se realizó un análisis bajo el modelo de efectos aleatorios. El metanálisis pone de manifiesto una mejora significativa en la calidad de vida de los pacientes que recibieron el tratamiento con ejercicio, ya que al comparar con quienes solo reciben el tratamiento estándar se observa una diferencia de 12,47 en el puntaje de calidad de vida evaluado con el FIQ y un intervalo de confianza que va de 5,71 hasta 19,23, lejos del valor de no diferencia (ver Figura 43).

Figura 43

Forest plot para un ensayo clínico de diferencia de medias. (Después de la intervención)

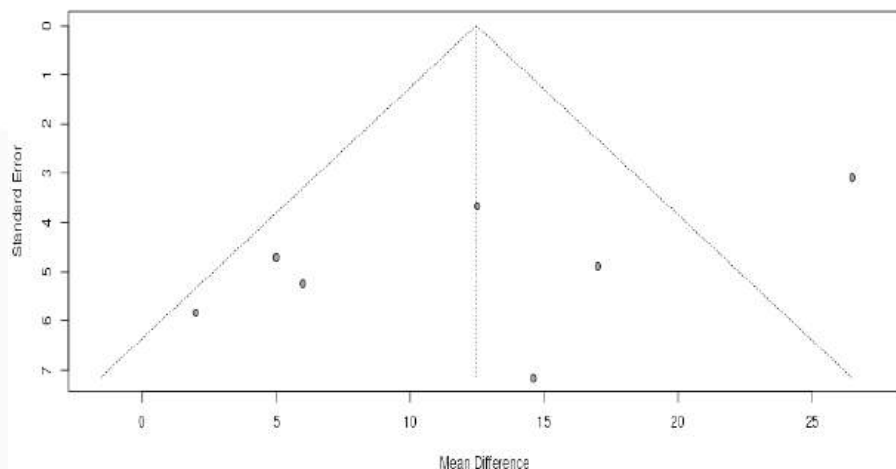


Nota. Elaboración propia.

El funnel plot pone de manifiesto ausencia de sesgo de publicación en tanto se observa que los estudios (representados por puntos) se distribuyen de manera homogénea a lado y lado del embudo (ver Figura 44).

Figura 44

Funnel plot para un ensayo clínico de diferencia de medias



Nota. Elaboración propia.

Conclusión: Este estudio puso al descubierto la mayor eficacia del ejercicio físico, en comparación con el tratamiento estándar para mejorar la calidad de vida de pacientes con fibromialgia.

Referencias

- Celentano, D. D., Szklo, M., & Farag, Y. M. K. (2024). *Gordis. Epidemiología* (7 Ed.). Elsevier
- Chen, B., Wang, C. C., Lee, K. H., Xia, J. C., & Luo, Z. (2023). Efficacy and safety of acupuncture for depression: A systematic review and meta-analysis [Eficacia y seguridad de la acupuntura para la depresión: una revisión sistemática y un metanálisis]. *Research in Nursing & Health*, 46(1), 48-67. <https://doi.org/10.1002/nur.22284>
- Chen, Y. S., Lee, H. F., Tsai, C. H., Hsu, Y. Y., Fang, C. J., Chen, C. J., Hung, Y. H., & Hu, F. W. (2022). Effect of Vitamin B2 supplementation on migraine prophylaxis: a systematic review and meta-analysis [Efecto de la suplementación con vitamina B2 en la profilaxis de la migraña: una revisión sistemática y metanálisis].

sis]. *Nutritional Neuroscience*, 25(9), 1801-1812. <https://doi.org/10.1080/1028415X.2021.1904542>

Cochrane Methods Bias. (2025). RoB 2: A revised Cochrane risk-of-bias tool for randomized trials [RoB 2: Una herramienta Cochrane revisada de riesgo de sesgo para ensayos aleatorios]. <https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials>

Didari, T., Mozaffari, S., Nikfar, S., & Abdollahi, M. (2015). Effectiveness of probiotics in irritable bowel syndrome: Updated systematic review with meta-analysis [Eficacia de los probióticos en el síndrome del intestino irritable: revisión sistemática actualizada con metanálisis]. *World Journal of Gastroenterology*, 21(10), 3072-3084. <https://doi.org/10.3748/wjg.v21.i10.3072>

Gérvás, J., & Pérez Andrés, C. (2008). Evaluación de las intervenciones en salud: la búsqueda del equilibrio entre la validez interna de los resultados y la validez externa de las conclusiones. *Revista Española de Salud Pública*, 82(6), 577-579. <https://scielo.isciii.es/pdf/resp/v82n6/editorial.pdf>

Higgins, J. P. T., & Green, S. (Ed.). (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [Manual Cochrane para Revisiones Sistemáticas de Intervenciones Versión 5.1.0]. The Cochrane Collaboration. www.cochrane-handbook.org

Huang, L., Yin, Y., Yang, L., Wang, C., Li, Y., & Zhou, Z. (2017). Comparison of Antibiotic Therapy and Appendectomy for Acute Uncomplicated Appendicitis in Children: A Meta-analysis [Comparación de la terapia antibiótica y la apendicectomía para la apendicitis aguda no complicada en niños: un metanálisis]. *JAMA Pediatrics*, 171(5), 426-434. <https://doi.org/10.1001/jamapediatrics.2017.0057>

Lazcano-Ponce, E., Salazar-Martínez, E., Gutiérrez-Castrellón, P., Angeles-Llerenas, A., Hernández-Garduño, A., & Viramontes, J.L. (2004). Ensayos clínicos aleatorizados: variantes, métodos de aleatorización, análisis, consideraciones éticas y regulación. *Salud Pública de México*, 46(6), 559-584. <https://www.scielosp.org/pdf/spm/2004.v46n6/559-584/es>

Sterne, J. A. C., Savovi, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li,

- T.,... Higgins, J. P. T. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials [RoB 2: una herramienta revisada para evaluar el riesgo de sesgo en ensayos aleatorios]. *BMJ*, 366, Article 14898. <https://doi.org/10.1136/bmj.14898>
- Umscheid, C. A., Margolis, D. J., & Grossman, C. E. (2011). Key concepts of clinical trials: a narrative review [Conceptos clave de los ensayos clínicos: una revisión narrativa]. *Postgraduate Medicine*, 123(5), 194-204. <https://doi.org/10.3810/pgm.2011.09.2475>
- Van der Mierden, S., Tsaïoun, K., Bleich, A., & Leenaars, C. H. C. (2019). Software tools for literature screening in systematic reviews in biomedical research [Herramientas de software para la selección de literatura en revisiones sistemáticas en investigación biomédica]. *ALTEX*, 36(3), 508-517. <https://doi.org/10.14573/altex.1902131>
- Weitzner, Z. N., Phan, J., Begashaw, M. M., Mak, S. S., Booth, M. S., Shekelle, P. G., Maggard-Gibbons, M., & Girgis, M. D. (2023). Endoscopic therapies for patients with obesity: a systematic review and meta-analysis [Terapias endoscópicas para pacientes con obesidad: una revisión sistemática y metanálisis]. *Surgical Endoscopy*, 37(11), 8166-8177. <https://doi.org/10.1007/s00464-023-10390-6>

Conclusiones y recomendaciones

Las revisiones sistemáticas y los metanálisis son un tipo de investigación que condensa y analiza la evidencia que se publica en torno a un tópico de interés. Por esta característica, esta modalidad de investigación suele ubicarse en la cúspide de la evidencia para orientar la práctica clínica y, en general, la toma de decisiones en el campo de la salud, de modo que son muy valoradas. Sumado a esto, tienen muchas ventajas operativas pues son de bajo costo, relativamente rápidas de realizar, no requieren de infraestructura compleja, ni de grandes grupos de investigación. Por tanto, se constituyen en una excelente alternativa para iniciarse en la investigación, realizar trabajos de grado en pregrado o posgrado, pero también para construir guías de práctica clínica que cambien conductas y generen profundas transformaciones en el campo de la salud.

Teniendo en cuenta las ventajas de esta modalidad de investigación, hay una proliferación de publicaciones de este tipo; cada cierto tiempo se proponen nuevas clasificaciones, subclasificaciones, nuevas técnicas de análisis y estrategias de búsqueda. En el espectro de la calidad, se encuentran estudios en un amplio rango que van desde excelentes, hasta verdaderamente discretos o incluso cuestionables. Esperamos que este libro haya aportado las bases para que los investigadores más jóvenes, interesados en emprender una investigación de este tipo, puedan conducirlas con alto rigor metodológico o, si deciden hacer uso de ellas en su práctica clínica, tengan los elementos generales para leerlas y juzgarlas aplicándoles unos criterios de calidad.

La inteligencia artificial (IA) es un recurso emergente que se está promoviendo con fuerza en distintos escenarios para conducir revisiones sistemáticas y metanálisis. Con este propósito, han surgido artículos como el de Bernard et al. (2025) y el de Ge et al. (2024) que describen la automatización de procesos como la selección de estudios, la extracción de datos, incluso la evaluación de la calidad metodológica, mediante la incorporación de herramientas de IA. Además, se han desarrollado plataformas cada vez más refinadas como *Searchrefiner*²⁹, *SWIFT-Review*³⁰, *Colandr*³¹, entre muchas otras

²⁹ <https://ielab.io/searchrefiner/>

³⁰ <https://www.sciome.com/swift-review/>

³¹ <https://www.colandrapp.com/signin>

que buscan mecanizar etapas catalogadas como engorrosas o dispendiosas para el investigador en búsqueda y tamización de artículos. No obstante, es importante que los investigadores (noveles o experimentados) eviten que la herramienta reemplace su juicio.

La aplicación de la IA en el ámbito de las revisiones sistemáticas, los metanálisis y, en general, en la investigación en salud, debe enfatizar la necesidad de proporcionarle el contexto suficiente a estas herramientas al momento de hacer preguntas, dar instrucciones y obtener respuestas. Darles contexto a estas herramientas significa que el investigador debe propiciarle la información a la IA, pensar en qué necesita el modelo para responder sus preguntas, qué elementos debe tener en consideración y cuáles sobran o debería omitir. Proceder de esta manera implica el desarrollo de habilidades técnicas y claridades conceptuales sobre el tema que se está trabajando, que en el caso de las revisiones sistemáticas y los metanálisis sobre los distintos tipos de estudio epidemiológicos, el lector pudo encontrar en este libro, desagregados en generalidades sobre cada tipo de estudio, sus variantes y sesgos; la estrategia de búsqueda, criterios de inclusión y exclusión de artículos, extracción de la información que se va a sistematizar, evaluación de la calidad metodológica y algunas consideraciones sobre el análisis de la información. Articular los conceptos y técnicas condensadas en este libro con las herramientas que ofrece la inteligencia artificial, facilitará a los investigadores el desarrollo de revisiones sistemáticas y metanálisis en el futuro.

Por el contrario, un uso irracional o descontextualizado de las herramientas de IA en la formulación y desarrollo de revisiones sistemáticas y metanálisis puede ocasionar problemas en la comprensión del método o una comprensión muy superficial, menor retención de información y, en general, una “desconexión cognitiva” del investigador con el tópico que está investigando. Estos problemas son opuestos a las competencias que busca desarrollar la investigación en general y la investigación formativa en particular, entre ellos, la capacidad de formular y resolver problemas, sintetizar y analizar información, incorporar el método de investigación, apropiarse de su objeto de estudio, entre otras. Lo expuesto nos lleva a desaconsejar el uso de esas herramientas en ausencia de una formación conceptual y metodológica básica como la descrita en cada uno de los capítulos de este libro.

Referencias

- Bernard, N., Sagawa, Y., Jr, Bier, N., Lihoreau, T., Pazart, L., & Tannou, T. (2025). Using artificial intelligence for systematic review: the example of elicitor [Utilizando inteligencia artificial para la revisión sistemática: el ejemplo de elicitor]. *BMC Medical Research Methodology*, 25(1), Article 75. <https://doi.org/10.1186/s12874-025-02528y>
- Ge, L., Agrawal, R., Singer, M., Kannapiran, P., De Castro Molina, J. A., Teow, K. L., Yap, C. W., & Abisheganaden, J. A. (2024). Leveraging artificial intelligence to enhance systematic reviews in health research: advanced tools and challenges [Aprovechar la inteligencia artificial para mejorar las revisiones sistemáticas en la investigación en salud: herramientas avanzadas y desafíos]. *Systematic Reviews*, 13(1), Article 269. <https://doi.org/10.1186/s13643-024-02682-2>

Herramientas para realizar revisiones sistemáticas y metanálisis

En esta sección se presentan diferentes herramientas, algunas de ellas gratuitas, para realizar revisiones sistemáticas y metanálisis, desde la formulación de la pregunta hasta el análisis de la información.

Formulación de la pregunta PICO

La herramienta PICOT Question Generator es un recurso en línea diseñado para ayudar a formular preguntas clínicas estructuradas, siguiendo el formato PICOT, el cual es ampliamente utilizado en medicina basada en evidencia.

Fuente: <https://picotquestion.com/>

Seleccionar el tipo de revisión

La herramienta *RIGHT Review* es un instrumento en línea de acceso abierto que orienta a investigadores nuevos en el campo de la síntesis de la literatura, sobre el tipo de revisión que debe realizar. Para eso, guía al investigador por una serie de preguntas sobre los objetivos de la revisión.

Fuente: <https://rightreview.knowledgetranslation.net/>

Gestión de las referencias y aplicación del protocolo

Rayyan: es una herramienta que ayuda a agilizar la selección inicial de títulos y resúmenes de los artículos identificados. La plataforma soporta la importación de referencias desde diversas bases de datos y gestores bibliográficos (como PubMed, EndNote, Mendeley, Zotero) en formatos como RIS, BibTeX y CSV. Una vez importadas las referencias, facilita la aplicación de criterios de inclusión y exclusión. Además, la aplicación permite que múltiples revisores trabajen simultáneamente en una revisión

Fuente: <https://www.rayyan.ai/>

RevMan: Software de la colaboración Cochrane para realizar revisiones sistemáticas

Fuente: <https://revman.cochrane.org/info>

Covidence: Es una plataforma en línea diseñada para gestionar revisiones sistemáticas de manera eficiente que ayuda con la importación de referencias desde gestores bibliográficos, el proceso de selección de estudios, extracción de datos y evaluación de calidad. Permite el trabajo de varios revisores. Tiene algunas funciones de acceso gratuito, pero el acceso completo requiere suscripción paga.

Fuente: <https://www.covidence.org/>

Análisis de la información

Software para metanálisis de pruebas diagnósticas

MetaDisc 2.0: Este es un software gratuito financiado por el Instituto Ramón y Cajal de Investigación Sanitaria para el análisis de información de metanálisis de pruebas diagnósticas.

Fuente: <https://www.irydis.org/es/innovacion/metadisc/>

Análisis de la información de estudios de prevalencia

Neyeloff, Fuchs, S.C. & Moreira desarrollaron una plantilla en Excel para el análisis de información de metanálisis de prevalencia. Todo el proceso de creación está descrito en un artículo publicado en BMC Research Notes y el acceso al texto completo se puede consultar en la siguiente fuente

Fuente: Neyeloff, J. L., Fuchs, S. C., & Moreira, L. B. (2012). Meta-analyses and forest plots using a microsoft excel spreadsheet: step-by-step guide focusing on descriptive data analysis [Metanálisis y diagramas de bosque con una hoja de cálculo de Microsoft Excel: guía paso a paso centrada en el análisis descriptivo de datos]. *BMC Res Notes*, 5, Article 52. <https://doi.org/10.1186/1756-0500-5-52>

Análisis de la información de estudios de casos y controles, cohortes, estudios de intervención

METAANALYSISONLINE: Es una plataforma web que permite realizar metanálisis de OR, RR y hazard ratios (HR) e incluso de medidas continuas. Esta plataforma ofrece opciones para generar gráficos de forest plot, Funnel plot y análisis de heterogeneidad. Además, permite seleccionar entre modelos de efectos fijos y aleatorios.

Fuente: <https://metaanalysisonline.com/>

OpenMeta[Analyst]: Software gratuito desarrollado en un proyecto financiado por AHRQ-Funded Digital Healthcare Research Projects, de Estados Unidos, para realizar diferentes tipos de metanálisis.

Fuente: <http://www.cebm.brown.edu/openmeta/#>

Meta-Essentials: Proporciona una serie de plantillas de Excel que permite hacer metanálisis de datos binarios y continuos.

Fuente: <https://www.erim.eur.nl/research-support/meta-essentials/download/>

Análisis de la información metanálisis en red

Metainsight: Herramienta gratuita, en línea, para analizar y visualizar metanálisis de redes.

Fuente: <https://crsu.shinyapps.io/MetaInsight/>

Herramientas para evaluar la calidad metodológica

Equatonetwork: Proporciona una biblioteca de guías para evaluar la calidad metodológica de diferentes tipos de estudio.

Fuente: <https://www.equator-network.org/>

JBICritical Appraisal Tools: Herramientas desarrolladas por el Joanna Briggs Institute para evaluar la calidad de aproximadamente

15 tipos de estudios diferentes.

Fuente: <https://jbi.global/critical-appraisal-tools>

Herramientas de evaluación de calidad del NHLBI: El NHLBI desarrolló un conjunto de herramientas de evaluación de calidad para ayudar a los revisores a centrarse en conceptos clave para la validez interna y externa de las investigaciones durante el proceso de revisión sistemática de la evidencia y con el fin de actualizar diferentes guías clínicas.

Fuente: <https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools>

Otros recursos

<http://systematicreviewtools.com/>

<https://training.cochrane.org/online-learning/core-software>

LUIS FELIPE HIGUITA GUTIÉRREZ

Es microbiólogo, magíster en Educación y Desarrollo Humano, doctor en Ciencias Sociales y profesor de pregrado y posgrado en la Facultad de Medicina de la Universidad Cooperativa de Colombia, sede Medellín. Ha impartido cursos de métodos de investigación en programas de microbiología, salud pública, especialidades médico-quirúrgicas y psicología de la salud. Cuenta con más de 60 publicaciones científicas indexadas en Scopus, entre ellas revisiones sistemáticas y metanálisis. Posee experiencia en el diseño y desarrollo de estudios epidemiológicos con enfoques cuantitativos, cualitativos y mixtos. Su labor académica se orienta a la formación investigativa, la evaluación crítica de la evidencia y el fortalecimiento de la investigación en ciencias de la salud.

JAIBERTH ANTONIO CARDONA ARIAS

Es microbiólogo, magíster en Epidemiología, magíster en Economía Aplicada, magíster en Filosofía y Religión, y doctor en Salud Pública. Es Profesor Titular de la Escuela de Microbiología de la Universidad de Antioquia, Medellín, Colombia, donde lidera el Grupo de Investigación en Medicina Transfusional y Terapia Celular (GIMETT). Cuenta con más de 140 publicaciones científicas indexadas en Scopus, entre ellas revisiones sistemáticas y metanálisis. Ha desarrollado una amplia trayectoria docente en programas de pregrado y posgrado, especialmente en métodos de investigación, epidemiología y ciencias de la salud. Tiene experiencia en la formulación y ejecución de investigaciones epidemiológicas desde perspectivas cuantitativas, cualitativas y mixtas. Su trabajo universitario se centra en la generación de nuevo conocimiento, la formación de nuevos investigadores y el análisis riguroso de la evidencia científica.



**UNIVERSIDAD
DE ANTIOQUIA**