



## Revisiones sistemáticas sin meta-análisis: un tutorial

**Julio Sánchez-Meca\***

Departamento de Psicología Básica y Metodología. Facultad de Psicología y Logopedia. Campus de Espinardo, Universidad de Murcia

Mail: [jsmeca@um.es](mailto:jsmeca@um.es)

ODCID: <https://orcid.org/0000-0002-8412-788X>

**José Antonio López-López**

Departamento de Psicología Básica y Metodología. Facultad de Psicología y Logopedia. Campus de Espinardo, Universidad de Murcia

Mail: [josealopezlopez@um.es](mailto:josealopezlopez@um.es)

ORCID: <https://orcid.org/0000-0001-8474-1823>

**María Rubio-Aparicio**

Departamento de Psicología Básica y Metodología. Facultad de Psicología y Logopedia. Campus de Espinardo, Universidad de Murcia

Mail: [mrubioaparicio@um.es](mailto:mrubioaparicio@um.es)

ORCID: <https://orcid.org/0000-0002-2599-4246>

### RESUMEN

Las revisiones sistemáticas (RS) y los meta-análisis (MA) constituyen un tipo de investigación fundamental para una adecuada acumulación del conocimiento en cualquier ciencia empírica. Si bien los MA permiten alcanzar conclusiones más sólidas sobre la pregunta de interés, no siempre se dan las condiciones adecuadas para su aplicación. En su defecto, pueden aplicarse RS. Se presentan las diferencias y similitudes entre las RS y los MA. Se presentan las fases en las que se lleva a cabo una RS sin MA: planteamiento de la pregunta, definición de los criterios de selección de los estudios, búsqueda de los estudios, extracción de la información, valoración del riesgo de sesgo, medida del resultado de los estudios, métodos de síntesis y redacción del informe. Se ofrece un tratamiento detallado de las dos características principales que diferencian una RS de un MA: la medida del resultado de los estudios (tamaños del efecto, niveles de probabilidad, dirección del efecto) y los métodos de síntesis (cuantitativa) alternativos a los típicos de un MA. A lo largo del tutorial se utiliza un ejemplo para ilustrar las fases de una RS. Finalmente, se ofrecen reflexiones sobre las ventajas y limitaciones de las RS en relación con los MA.

*Palabras clave:* Revisión sistemática, meta-análisis, síntesis de la investigación, métodos de síntesis

### Systematic reviews without Meta-analysis: A tutorial

#### ABSTRACT

Systematic reviews (SR) and meta-analyses (MA) are essential research methodologies for the rigorous accumulation of knowledge in any empirical science. While MA enable more robust conclusions regarding the question of interest, the necessary conditions for their implementation are not always met. In such cases, SR can be conducted as an alternative. This tutorial examines the key similarities and differences between SR and MA, with a focus on the methodological processes involved in conducting an SR without MA. The phases discussed include formulation of the research question, definition of inclusion and exclusion criteria, literature search, data extraction, risk of bias assessment, outcome measurement, synthesis methods, and report writing. Special attention is given to two core features that distinguish SR from MA: the measurement of study outcomes (e.g., effect sizes, probability levels, di-

\*Autor de correspondencia: [jsmeca@um.es](mailto:jsmeca@um.es)

Recibido: 30/09/2025 – Aceptado: 31/10/2025

ISSN: 0210-2773

DOI: <https://doi.org/10.17811/rife.23622>



rection of effects) and quantitative synthesis methods alternative to those typically employed in MA. An example is used throughout to illustrate each phase of the SR process. Finally, the tutorial discusses on the strengths and limitations of SR in comparison to MA.

*Keywords:* Systematic review, meta-analysis, research synthesis, synthesis methods.

## 1. Introducción

Las revisiones sistemáticas (RS) se han convertido en una metodología de investigación imprescindible para garantizar una correcta acumulación del conocimiento científico en cualquier disciplina empírica. Los meta-análisis (MA) son un tipo especial de RS que también se están aplicando ya de forma rutinaria para sintetizar los resultados de un conjunto de estudios sobre una misma temática. Si bien los términos RS y MA están interrelacionados, no hacen referencia a un mismo tipo de revisión de la investigación. Es importante saber distinguir entre qué es una RS, qué es un MA y cuándo es apropiado utilizar uno u otro tipo de revisión (Cooper, 2016; Higgins, Thomas et al., 2024).

El uso de RS y MA no sólo es frecuente en cualquier disciplina empírica, sino también en el contexto de la formación universitaria. Los currículos de Grado y Máster universitarios incorporan asignaturas del tipo 'Trabajo Fin de Carrera' en las que el estudiantado tiene la opción de llevar a cabo una investigación. Tanto en el Trabajo Fin de Grado como en el Trabajo Fin de Máster, es muy frecuente encontrar investigaciones que afirman ser 'una revisión sistemática'. Es por ello que, tanto el estudiantado como el profesorado, debe tener una buena fundamentación sobre qué es y cómo se hace una RS y, en su caso, un MA.

El propósito de este artículo es clarificar las similitudes y diferencias entre las RS y los MA, describir cómo se lleva a cabo una RS sin MA y cómo se diferencian estos dos tipos de revisión respecto de otras. El foco de este artículo está en las RS, no en los MA. La razón de excluir del foco los MA es porque ya se publicó en esta misma revista un artículo sobre qué son y cómo se hacen los MA (Sánchez-Meca, 2010). También se pueden consultar otras referencias para profundizar en los MA (Botella y Sánchez-Meca, 2015; Sánchez-Meca, 2022; Sánchez-Meca et al., 2011; Siddaway et al., 2019).

En lo que sigue, se presentan en primer lugar definiciones de RS y MA y bajo qué condiciones es preferible uno u otro tipo de revisión. A continuación, se hace un recorrido a través de las fases que se siguen en la práctica para llevar a cabo una RS, para lo cual se utilizan ejemplos reales para ilustrar las diversas tomas de decisión que el revisor debe tomar a lo largo del proceso. Se hace especial hincapié en los diferentes modos de definir el 'resultado' de cada estudio y cómo se pueden sintetizar esos resultados sin llevar a cabo un MA. Se finaliza con unas reflexiones sobre la utilidad de las RS y sobre las guías y checklists existentes para orientar a los revisores sobre cómo redactar este tipo de investigación.

## 2. Revisión sistemática y meta-análisis

Una *revisión sistemática* (RS) es una investigación en la que se plantea una pregunta con total claridad y, para dar respuesta a ella, se definen unos criterios de selección de los estudios, se lleva a cabo una búsqueda metódica de dichos estudios, se extraen las características y los resultados de los estudios y se realiza una síntesis de dichos resultados que permita dar respuesta a la/s pregunta/s planteada/s en la RS (Borenstein et al., 2021; Cooper et al., 2019).

Un *meta-análisis* (MA) es un tipo especial de RS en el que, además de todo lo dicho hasta aquí, se cumplen dos condiciones: (a) de cada estudio se ha podido obtener un índice del tamaño del efecto, siendo el mismo en todos ellos (e.g., se ha podido calcular, al menos, una diferencia media tipificada en cada estudio) y (b) se aplican métodos estadísticos para sintetizar los tamaños del efecto teniendo en cuenta la variabilidad debida al error de muestreo de cada tamaño del efecto. Esos métodos estadísticos pueden consistir, por ejemplo, en construir *forest plots*, obtener el efecto medio con su intervalo de confianza, comprobar si el sesgo de publicación puede ser una amenaza contra la validez de los resultados del MA, evaluar la heterogeneidad de los tamaños del efecto (e.g., con el cálculo del estadístico  $Q$ , el índice  $I^2$  o la construcción de intervalos de predicción) y, caso de que exista heterogeneidad, analizar el influjo de variables moderadoras sobre los tamaños del efecto. En una RS que no es un MA se pueden utilizar otros métodos, más o menos cuantitativos, para extraer el resultado de cada estudio empírico y se pueden utilizar otros métodos de síntesis de esos resultados diferentes a los típicos de un MA, como se verá más adelante.

Es importante apuntar que la definición de MA planteada aquí es propia de las Ciencias Sociales (Psicología, Educación, Economía, Criminología, Trabajo Social, etc.). En las Ciencias de la Salud (Medicina, Enfermería, Fisioterapia, Odontología, etc.) el concepto de MA es diferente. En ese ámbito, el MA se define como la parte estadística de una RS, es decir, constituye la aplicación de los métodos de síntesis estadística dentro de una RS. Desde esta perspectiva, el MA se concibe, no como un tipo especial de RS, sino como una fase de una RS. Esta diferencia de concepción hace que los títulos de las RS y MA sean algo diferentes dependiendo de la disciplina en la que se esté llevando a cabo dicha revisión.

Un punto en común crucial que tienen las RS y los MA es el tipo de estudios que incluyen, ya que tienen que ser obligatoriamente estudios empíricos. En una RS no caben otras revisiones previas, tanto sistemáticas como narrativas, ni trabajos teóricos, ni cartas al editor, ni reseñas de libros. Las RS y los MA se nutren exclusivamente de estudios empíricos. Existen otros tipos de revisión que sí permiten la inclusión de otros estudios además de los empíricos (Booth et al., 2016). Remitimos al lector a otras fuentes para

profundizar en esos otros tipos de revisión que no son el propósito de este artículo, tales como las *revisiones de alcance* (*scoping reviews*; Munn et al., 2022), las *revisiones narrativas* (*narrative reviews*; Lubbe et al., 2020), las *revisiones metodológicas* (*methodological reviews*; Grant y Booth, 2009), las *revisiones rápidas* (*rapid reviews*; Garritty et al., 2025) o las *meta-revisiones de RS* (*umbrella reviews*; López-López et al., 2022).

Son varias las razones por las que puede no ser aconsejable llevar a cabo un MA y, en su lugar, hacer una RS. Cuando los estudios empíricos seleccionados exhiben una excesiva heterogeneidad en cómo se han llevado a cabo (i.e., heterogeneidad metodológica), o en las poblaciones objeto de estudio, en las variables dependientes medidas, en los tipos de programas o intervenciones de interés, etc. (i.e., heterogeneidad sustantiva), entonces puede no ser apropiado hacer un MA. Otra razón puede estar en que no se ha logrado obtener índices del tamaño del efecto de los estudios empíricos, por ejemplo, debido a falta de información estadística relevante en los estudios. O bien que, aunque sea posible obtener índices del tamaño del efecto, éstos son diferentes de unos estudios a otros (e.g., se han obtenido diferencias medias estandarizadas en algunos estudios, diferencias de proporciones en otros, coeficientes de correlación en otros, etc.). La presencia de una excesiva heterogeneidad metodológica y/o sustantiva en los estudios empíricos es, pues, un inconveniente importante para poder llevar a cabo un MA. En su defecto, se puede hacer una RS. La principal razón que se suele aducir para hacer una RS en lugar de un MA es la excesiva heterogeneidad de los estudios (Cumpston et al., 2023).

### 3. Fases de una RS

Las RS y los MA se llevan a cabo siguiendo las mismas fases o etapas (Cooper, 2016; Higgins, Thomas et al., 2024; Sánchez-Meca, 2022): (1) Formulación de la pregunta, (2) Criterios de selección de los estudios, (3) Búsqueda de los estudios, (4) Extracción de la información, (5) Valoración de la calidad de los estudios, (6) Obtención de los resultados, (7) Métodos de síntesis y (8) Redacción de la RS. Las principales diferencias entre una RS y un MA se plantean en las etapas (6) y (7), por lo que se hará especial hincapié en ellas (Cook y West, 2012; Plakkal, 2023).

#### 3.1. Formulación de la pregunta

Tanto en una RS como en un MA se debe plantear una o varias preguntas, u objetivos, con total claridad. Ello implica definir conceptos, constructos, variables, procedimientos experimentales, poblaciones diana, variables de resultado, etc., que tengan que ver con las preguntas. Una razón para aplicar una RS y no un MA puede ser que se espera encontrar pocos estudios debido a que la temática es incipiente y está poco investigada (aunque también existe la posibilidad de replantear esta decisión si finalmente se encuentran estudios con información cuantitativa suficiente para realizar un MA). Otra razón puede ser que los estudios sobre esa temática hayan aplicado diseños muy diversos, tales como diseños experimentales (e.g., ensayos controlados aleatorizados), cuasi-experimentales (e.g., ensayos controlados no aleatorizados), u observacionales (e.g., diseños de cohortes y/o de casos y controles), lo cual dificulta la realización de un MA. Otra razón puede estar en una excesiva diversidad en las variables dependientes o de resultado (*outcomes*) medidas en los diferentes estudios. Bajo estas circunstancias, la aplicación de métodos de síntesis estadística típicas de los MA podría verse comprometida. Por ejemplo, Rosales-Ricardo y Cáceres-Manzano (2024) realizaron una RS sobre la relación entre ejercicio físico y rendimiento académico en estudiantes universitarios. Dado que en sus criterios de selección aceptaron estudios con diseños muy diferentes (estudios experimentales, cuasi-experimentales, analíticos, observacionales, correlacionales), esa excesiva heterogeneidad metodológica les llevó a realizar una RS y no un MA.

#### 3.2 Criterios de selección de los estudios

Una vez planteada la pregunta, la siguiente fase de una RS, y de un MA, es definir los criterios de selección de los estudios empíricos que se pretende buscar. Cuando la RS tiene como objetivo comprobar la eficacia de una o varias intervenciones, resulta aconsejable utilizar el acrónimo PICOS para no olvidar ninguno de los aspectos relevantes de tales objetivos (Higgins, Thomas et al., 2024):

- (a) '*Participants*': se debe especificar cuál es la población o poblaciones objeto de interés. En Ciencias Sociales y de la Salud, las poblaciones de interés suelen ser personas con ciertas características (e.g., población comunitaria, población clínica con una determinada patología o trastorno, población infantil, etc.). Si el propósito de una RS involucra a diferentes poblaciones, esa diversidad puede desaconsejar la realización de un MA y, en su lugar, hacer una RS.
- (b) '*Interventions*': se debe indicar cuál o cuáles son los programas, tratamientos o intervenciones de interés. Una excesiva variedad en los tipos de intervenciones puede desaconsejar la aplicación de un MA y, en su defecto, hacer una RS. Esta recomendación se hace especialmente relevante cuando se encuentran pocos estudios de cada tipo de intervención.
- (c) '*Comparison groups*': los grupos de comparación se refieren a qué tipos de grupos se enfrentan los grupos experimentales de interés en cada estudio empírico. Los grupos de comparación pueden ser grupos de control inactivos (e.g., lista de espera), controles activos (e.g., placebo) o grupos que incorporan alguna intervención activa alternativa a la del grupo experimental. Una excesiva variedad en los grupos de comparación, junto con un número escaso de estudios, puede llevar a la conveniencia de realizar una RS y no un MA.

- (d) '*Outcomes*': se debe especificar qué variables dependientes o de resultado son de interés. Si la RS está interesada en resultados muy diversos (e.g., reducción de la ansiedad, de la depresión, mejora de la autoestima, mejora de la calidad de vida, mejora del bienestar subjetivo, etc.), entonces esa heterogeneidad en los tipos de resultado puede desaconsejar la realización de un MA y, en su lugar, hacer una RS.
- (e) '*Study designs*': se debe indicar qué diseños de los estudios son aceptables para la RS. Una excesiva variedad de diseños puede dificultar la realización de un MA y, en su lugar, llevar a cabo una RS. Por ejemplo, si los estudios aplican diseños tan diversos como diseños experimentales, cuasi-experimentales, de cohortes, de casos y controles, transversales, etc., esa heterogeneidad metodológica puede ser excesiva para llevar a cabo un MA.

Si la RS no tiene como objetivo analizar la eficacia de intervenciones, sino otras preguntas, tales como investigar la existencia de factores de riesgo o de protección de algún trastorno, enfermedad o problema clínico, educativo, social, psicológico, etc., puede utilizarse el acrónimo PECOS en lugar de PICOS para ayudar a fijar los criterios de selección. La única diferencia entre ambos está en la letra 'E', '*Exposure factor*', que sustituye a la letra 'I'. En este caso, se debe especificar cuál o cuáles son los factores de exposición (de riesgo y/o de protección) objeto de interés. Una excesiva variabilidad en dichos factores puede aconsejar la realización de una RS en lugar de un MA, especialmente si son pocos los estudios representativos de cada factor de exposición.

Independientemente de cuál sea la pregunta de interés, deben considerarse otros criterios de selección además de los ya comentados. Por ejemplo, la franja temporal de los estudios que se desea localizar. Otro criterio a tener en cuenta es si se van a aceptar sólo estudios publicados o también estudios no publicados. Y otro criterio adicional es en qué idiomas tiene que estar escrito el estudio para que pueda ser incluido en la RS (e.g., inglés, español, etc.).

En la RS de Rosales-Ricardo y Cáceres-Manzano (2024) sobre la relación entre ejercicio físico y rendimiento académico en estudiantes universitarios, los criterios de selección de los estudios fueron: (a) artículos publicados en los últimos 5 años, (b) estudios originales que aplicaron diferentes diseños (experimentales de diversa índole, correlacionales, analíticos y observacionales), (c) estudios que investigaron el efecto de cualquier tipo de ejercicio físico, y (d) la población diana incluyó a estudiantes universitarios de cualquier carrera y de ambos sexos. La excesiva heterogeneidad en los criterios de selección fue un importante condicionante para llevar a cabo una RS en lugar de un MA, además de la restricción temporal del criterio (a). Este último criterio puede ser apropiado para trabajos de estudiantes universitarios con el fin de reducir el volumen de información, y ocasionalmente también será válido en publicaciones científicas si existe una justificación adecuada para ello.

### 3.3 Búsqueda de los estudios

Una vez fijados los criterios de selección de los estudios, la siguiente fase consiste en intentar localizarlos. Para ello se aplican procedimientos de búsqueda sistemática, de entre los que no puede faltar la consulta de bases electrónicas mediante una adecuada combinación de palabras-clave y conectores. Las RS y los MA no difieren en el modo de llevar a cabo esta fase de búsqueda de los estudios, por lo que remitimos al lector a otras fuentes que describen los detalles de esta fase (cf., e.g., Cooper et al., 2019; Botella y Caperos, 2019, cap. 8).

Conviene precisar que siempre es recomendable utilizar diferentes procedimientos formales e informales de búsqueda. Además de las búsquedas electrónicas en las bases de datos, otros procedimientos formales de búsqueda de estudios publicados consisten en revisar de forma más específica aquellas revistas que sean prolíficas publicando estudios objeto de la RS, o revisar las referencias de estudios ya seleccionados para la RS, por si pudieran citar otros estudios seleccionables. Otra fuente formal consiste en revisar las referencias de los estudios incluidos en RS o MA previos sobre el mismo tema o similar. Los procedimientos informales de búsqueda de estudios están especialmente indicados si se pretende incluir estudios no publicados. Por ejemplo, se pueden consultar las actas de congresos relacionados con el tema, enviar mensajes a listas de distribución de asociaciones científicas solicitando información sobre estudios no publicados, o también enviar mensajes a autores que se sabe suelen realizar investigaciones sobre el tema de interés.

El proceso de búsqueda, localización, cribado y selección de los estudios debe ilustrarse de forma esquemática mediante un diagrama de flujo, como el PRISMA 2020 (Page et al., 2021). Puede descargarse en el sitio web de este grupo: <https://www.prisma-statement.org/prisma-2020-flow-diagram>. La Figura 1 muestra una versión en español de dicho diagrama.

Por ejemplo, en la RS de Rosales-Ricardo y Cáceres-Manzano (2024) la búsqueda de estudios se llevó a cabo mediante la consulta de las bases electrónicas Web of Science (Core Collection), Scopus y Medline a través de Pubmed. Los descriptores utilizados en las búsquedas fueron extraídos del '*Medical Subject Headings*' (MESH): "*Exercise*" AND "*Academic Performance*" AND "*University Students*". Así mismo, estos autores ilustraron el proceso de búsqueda, cribado y selección de los estudios mediante un diagrama de flujo PRISMA (Rosales-Ricardo y Cáceres-Manzano, 2024, p. 177). El proceso de búsqueda les permitió seleccionar 11 estudios empíricos para la RS, de los que siete estudios aplicaron un diseño transversal, dos un diseño correlacional, uno un diseño correlacional y predictivo y uno un diseño cuasi-experimental.

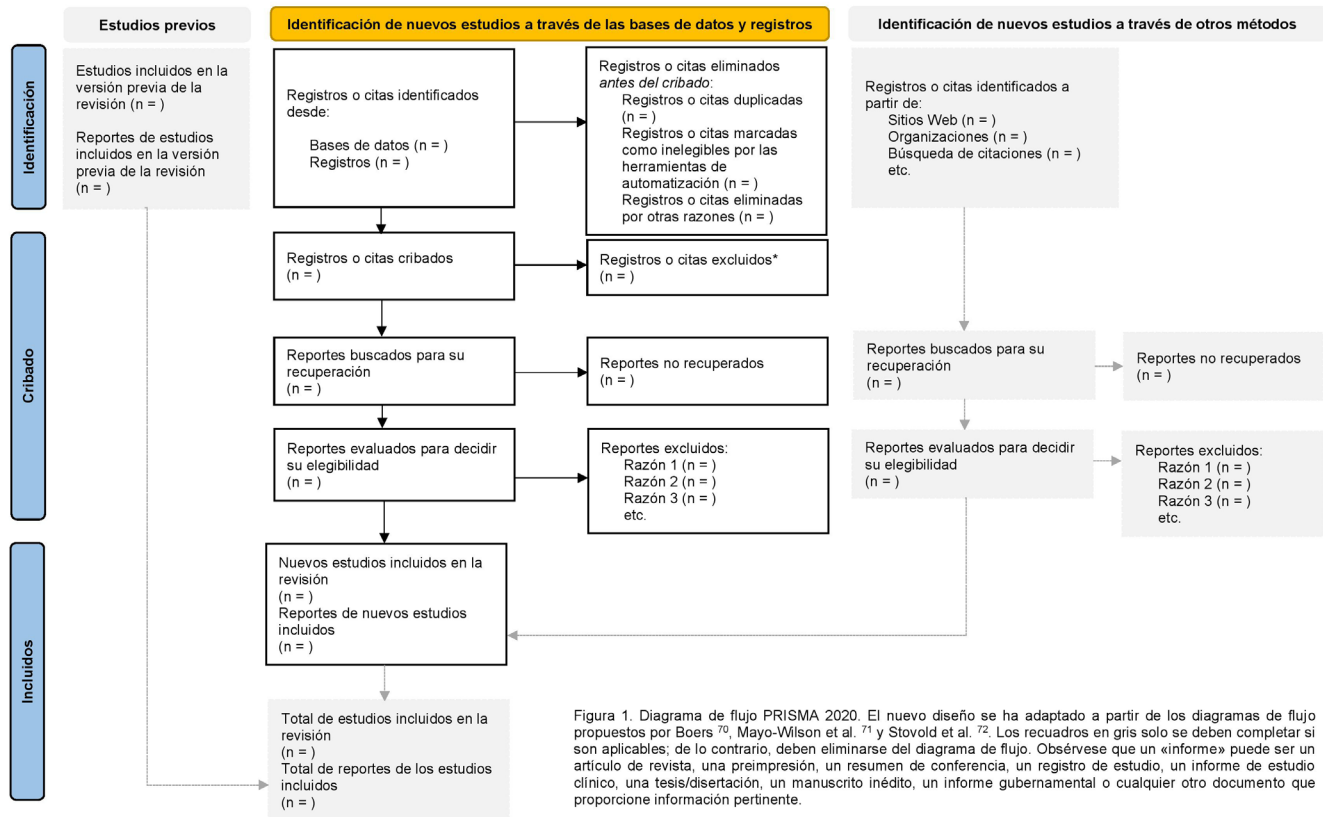


Figura 1. Diagrama de flujo PRISMA 2020

### 3.4 Extracción de la información

Una vez seleccionados los estudios, la siguiente fase consiste en extraer la información relevante de los estudios. El proceso de extracción de información involucra dos tipos de información: las características de los estudios y los resultados de los estudios que tienen que ver con la/s pregunta/s de interés de la RS. En este epígrafe se describe cómo se extrae la información relativa a las características de los estudios. La extracción de los resultados de los estudios se verá en el epígrafe 3.6.

Si bien las características de los estudios que se deben extraer dependen de cada RS, éstas pueden clasificarse en tres categorías: características metodológicas, sustantivas y extrínsecas (Cooper et al., 2019). Las *características metodológicas* tienen que ver con cómo se ha llevado a cabo la investigación. Son ejemplos de características metodológicas el diseño (e.g., experimental, cuasi-experimental, de cohortes, de casos y controles, transversal, etc.), el tipo de grupo de control (e.g., inactivo vs. activo), el enmascaramiento del evaluador, el uso de análisis por intención de tratar, el control de variables de confundido, las ocasiones de medida (e.g., pretest, posttest, seguimientos), etc.

Las *características sustantivas* son las que tienen que ver con los objetivos propios de la RS. Éstas pueden, a su vez, agruparse en tres categorías. En primer lugar, cabe mencionar las *características de los participantes*, que implican aquéllas que tienen que ver con la población o poblaciones objeto de estudio. Cabe mencionar en esta categoría características tales como la población de referencia (e.g., comunitaria, clínica, análoga), la edad media de la muestra, la distribución por género de la muestra (e.g., porcentaje de mujeres), su distribución étnica, etc. En segundo lugar, las *características de los tratamientos* objeto de interés (sólo en el caso de que el objetivo de la RS sea sobre la eficacia de tratamientos), tales como el tipo de modelo de intervención (e.g., intervención psicológica, educativa, médica, psicoeducativa, etc.), la duración del tratamiento o programa (e.g., número de semanas), la intensidad del tratamiento (e.g., número de horas por semana), el modo de tratamiento (e.g., individual vs. grupal), etc. Y las *características contextuales* se refieren al contexto temporal (e.g., año del estudio) y geográfico (e.g., país del estudio).

Un último bloque se conoce como *características extrínsecas*, así denominadas porque son características de los estudios que no tienen que ver con el objeto de la RS ni con su metodología, pero que podrían afectar a los resultados de los estudios. En este grupo cabe mencionar el estatus de publicación del estudio (i.e., publicado vs. no publicado). Esta característica es importante porque muchos campos de investigación sufren el fenómeno conocido como *sesgo de publicación*, según el cual es más frecuente publicar estudios con resultados estadísticamente significativos (i.e.,  $p < 0,05$ ) que estudios que no alcanzan la significación estadística. Otra característica extrínseca es

la formación del investigador principal del estudio (e.g., psicólogo vs. psiquiatra; biólogo vs. veterinario; educador social vs. trabajador social, etc.) debido al posible efecto corporativista que podría darse en algunos campos de investigación, o bien a que en diferentes disciplinas los procedimientos y metodologías pueden definirse u operacionalizarse de formas diferentes. La fuente de financiación del estudio (e.g., pública vs. privada), la declaración de posible conflicto de intereses de los autores del estudio o el hecho de que se hubiera publicado previamente, o no, el protocolo del estudio, son también características extrínsecas que pueden ser de interés en una RS.

La extracción de todas estas características debe hacerse de la forma más sistemática y objetiva posible y garantizando su fiabilidad. Es por ello que toda RS debe incorporar un estudio de la fiabilidad del proceso de registro de las características de los estudios. Para ello, al menos dos codificadores deben extraer la información de los estudios de forma independiente y, posteriormente, comprobar la concordancia en sus registros. El análisis de la fiabilidad debe reportarse aportando índices de acuerdo inter-codificadores, tales como coeficientes kappa de Cohen, para las características categóricas, y correlaciones intra-clase, para las continuas. Las incongruencias se deben resolver por consenso.

En la RS de Rosales-Ricardo y Cáceres-Manzano (2024) sobre la relación entre ejercicio físico y rendimiento académico en universitarios, se registraron las siguientes características de los estudios: año de publicación, país, tamaño muestral, edad media de la muestra, carrera universitaria, tipo de diseño e instrumento utilizado para medir la cantidad de ejercicio físico.

### 3.5 Valoración de la calidad de los estudios

En el proceso de extracción de las características de los estudios no debe faltar la aplicación de alguna escala o checklist que permita valorar la calidad metodológica de los estudios, o bien, la posible presencia de sesgo en las estimaciones de los efectos. Se han propuesto numerosas escalas y checklists para valorar la calidad de los estudios empíricos. Buena parte de los ítems que suelen conformar estas escalas y checklists ya se han mencionado en el epígrafe anterior al referir las características metodológicas de los estudios.

De especial interés es el checklist de valoración del riesgo de sesgo elaborado por la Colaboración Cochrane, RoB 2 (Higgins, Savović et al., 2024; <https://www.riskofbias.info/>) para estudios experimentales o manipulativos (i.e., estudios sobre evaluación de tratamientos en los que los participantes se asignaron aleatoriamente a las diferentes condiciones o grupos), o la versión de este checklist para estudios sin asignación aleatoria (ROBINS-I; Sterne et al., 2024).

Cuando una RS incluye estudios con diseños que no son experimentales (e.g., cuasi-experimentales, correlacionales, observacionales), entonces es preferible recurrir a otras escalas o checklists mejor adaptados a estas circunstancias. Una de esas escalas es la *Newcastle-Ottawa Scale* (NOS; Wells et al., 2000). La escala Newcastle-Ottawa tiene la ventaja de que presenta diferentes versiones para su aplicación dependiendo de que los estudios hayan aplicado diseños de cohortes, de casos y controles o sean meros diseños transversales. En su RS, Rosales-Ricardo y Cáceres-Manzano (2024) aplicaron la escala Newcastle-Ottawa para valorar la calidad metodológica de cada uno de los 11 estudios seleccionados (cf. la Tabla 1, p. 177, de su artículo).

Del mismo modo que en el proceso de extracción de las características de los estudios, la valoración de la calidad metodológica de los estudios también debe someterse a un análisis de la fiabilidad en la aplicación de la escala o checklist en cuestión, mediante el cálculo de índices de acuerdo inter-codificadores.

### 3.6 Obtención de los resultados

La extracción de las características de los estudios (expuesta en el epígrafe 3.4), la valoración de la calidad metodológica (expuesta en el epígrafe 3.5) y la extracción de los resultados de los estudios, que es el objeto de este epígrafe, se llevan a cabo al mismo tiempo. Su separación en tres epígrafes diferentes obedece a razones didácticas. La extracción de los resultados de cada estudio incluido en la RS consiste en obtener la información, cuantitativa o cualitativa, relevante a la/s pregunta/s de interés de esa RS. Si el objetivo de una RS es, por ejemplo, valorar la eficacia de programas de reeducación de la lecto-escritura en escolares de una determinada franja de edad, el resultado de cada estudio relativo a ese objetivo debe ser la información (estadística o narrativa) relativa a los logros alcanzados con el programa de reeducación en cuestión.

En una RS sin MA el resultado de cada estudio se puede definir de diferentes formas, dependiendo de las características de los estudios y de la información aportada en los mismos. Esos resultados pueden ser de naturaleza estadística (e.g., medias, correlaciones), numérica (e.g., recuento de casos) o cualitativa (e.g., narrativa de los resultados). A continuación, se presentan los diferentes modos de definir el resultado de cada estudio en una RS, en orden de mayor a menor grado de cuantificación.

Al igual que en un MA, una RS puede cuantificar el resultado de cada estudio mediante el cálculo de índices del tamaño del efecto (e.g., diferencia de medias, diferencia media tipificada, coeficiente de correlación, odds ratio, etc.). La diferencia entre una RS y un MA está en cómo se maneja esa información estadística. Mientras que en un MA esos tamaños del efecto se someten a análisis estadísticos, es decir, a procesos de síntesis estadística teniendo en cuenta la variabilidad muestral de dichos tamaños del efecto (e.g., cálculo del efecto medio, análisis de la heterogeneidad, análisis del sesgo de publicación, análisis de moderadores, etc.), en una RS se pueden llevar a cabo síntesis estadísticas más 'rudimentarias', al no implicar la variabilidad muestral de los tamaños del efecto, como se verá en el siguiente epígrafe. Las razones que pueden llevar a los investigadores a llevar a cabo una RS y no un MA, a pesar de haber obtenido tamaños del efecto de los estudios, pueden ser varias. Una puede ser que no se haya podido obtener un mismo

índice del tamaño del efecto de todos los estudios (e.g., se han obtenido diferencias de medias en algunos estudios y correlaciones en otros). Otra razón puede ser que, aunque se haya podido obtener el mismo índice del tamaño del efecto en todos los estudios (e.g., diferencias medias tipificadas), la excesiva heterogeneidad metodológica y sustantiva entre los estudios desaconseja llevar a cabo la síntesis estadística propia de un MA. Una tercera razón, que puede ocurrir al mismo tiempo que las anteriores, es la escasez de estudios (aunque en disciplinas como la Medicina se llevan a cabo MA con pocos estudios habitualmente, aunque con tamaños muestrales elevados). Por tanto, el uso de índices del tamaño del efecto no es exclusivo de los MA. En una RS se pueden aportar tales índices, pero sin realizar la síntesis estadística típica de un MA (McKenzie y Brennan, 2021).

Para ilustrar con un ejemplo los diferentes modos de operativizar el resultado de cada estudio, en la Tabla 1 se presentan los resultados de la RS de Rosales-Ricardo y Cáceres-Manzano (2024) sobre la relación entre ejercicio físico y rendimiento académico en estudiantes universitarios. En esta RS se incluyeron 11 estudios de diferentes países. En dicha tabla se muestran algunas de las características de los estudios, así como diferentes formas de registrar el resultado de cada estudio. En concreto, en esta RS se reportaron algunas correlaciones que cuantificaron el sentido y la fuerza de la relación entre ejercicio físico y rendimiento académico. En la Tabla 1 la columna etiquetada ' $r$ ' contiene dichas correlaciones. Estas correlaciones están calculadas de forma que valores de signo positivo indicaron una relación positiva entre ejercicio físico y rendimiento académico, y viceversa. La tabla también incluye los errores típicos de dichas correlaciones, lo cual permitiría construir intervalos de confianza (e.g., al 95%) para cada correlación. Por tanto, esta RS podría haber utilizado estas correlaciones como tamaños del efecto de los estudios individuales para ilustrar la relación entre ejercicio físico y rendimiento académico en universitarios.

El mejor modo de operativizar y cuantificar el resultado de un estudio es mediante el cálculo de tamaños del efecto. Cuando los estudios no reportan los tamaños del efecto ni la información estadística necesaria para calcularlos, la RS tiene que recurrir a otros métodos para registrar el resultado de cada estudio. Una opción factible es la que se da cuando los estudios, aunque no reportan los tamaños del efecto, sí aportan el valor de probabilidad asociado al resultado del contraste de hipótesis que analizó la verosimilitud de la hipótesis de interés. En este caso, es importante tener en cuenta que los valores  $p$  que se reportan habitualmente en los estudios son valores de probabilidad 'bilaterales' (i.e., resultantes de resolver un contraste de hipótesis bilateral), ya que ése es el tipo de contraste que los programas estadísticos suelen reportar. Para su uso en el contexto de una RS, esos valores  $p$  bilaterales tienen que transformarse en valores  $p$  unilaterales (cf., e.g., Borenstein et al., 2021, cap. 41). En el pie de la Tabla 1 se presentan las fórmulas que tienen que utilizarse para realizar esa conversión. En la RS de Rosales-Ricardo y Cáceres-Manzano (2024), las columnas de la Tabla 1 etiquetadas como ' $p$ (Bilateral)' y ' $p$ (Unilateral)' contienen estos datos. La inspección de los valores  $p$ (Unilateral) permite comprobar si los resultados fueron muy favorables a la hipótesis de interés (i.e., valores de probabilidad muy bajos) o no (valores  $p$  altos).

Si los estudios tampoco aportan los valores de probabilidad, pero sí indican cuál fue la dirección del efecto investigado, entonces es posible cuantificar el resultado de cada estudio mediante una simple dicotomización: resultado favorable a la hipótesis versus resultado no favorable. Esta dicotomización permite posteriormente ofrecer una síntesis cuantitativa sobre si la mayoría de los estudios favorecieron, o no, dicha hipótesis (Borenstein et al., 2021). En la RS de Rosales-Ricardo y Cáceres-Manzano (2024), cuyos datos figuran en la Tabla 1, la columna etiquetada 'Dirección' indica si la correlación entre ejercicio físico y rendimiento académico fue positiva o negativa. La inspección de esta columna permite comprobar si la mayoría de los estudios obtuvieron, o no, un resultado favorable a la hipótesis de que el ejercicio físico está positivamente correlacionado con el rendimiento académico.

Finalmente, si los estudios no aportan ninguna de la información comentada (tamaños del efecto, valores  $p$ , dirección del efecto), entonces la RS puede recurrir al reporte del resultado de cada estudio de forma narrativa. Así, en la RS de Rosales-Ricardo y Cáceres-Manzano (2024, pp. 179-182) la Tabla 2 de dicho artículo contiene una descripción cualitativa de los resultados de cada estudio. Este método de reporte de los resultados es el menos aceptable o recomendable, ya que, debido a su fuerte componente de subjetividad, es propenso a sufrir sesgos de interpretación (Campbell et al., 2019).

**Tabla 1**

*Datos e información extraída de los estudios de la RS de Rosales-Ricardo y Cáceres-Manzano (2024).*

Elaboración propia a partir de la información contenida en Rosales-Ricardo y Cáceres-Manzano, 2024, así como mediante consulta directa de algunos de los estudios incluidos en esta RS.

Estudio	Año	País	Diseño	$N$	$r$	$ET$	Dirección	$p$ (Bilateral)	$p$ (Unilateral)
Alzhami et al (2021)	2021	Arabia Saudí	Transversal	379	0,048	0,051	Positiva	0,350827698	0,175413849
Bayramlar et al (2022)	2022	Turquía	Transversal	65	0,025	0,126	Positiva	0,843249874	0,421624937
Broadus et al (2021)	2021	EEUU	Correlacional	581	0,410	0,035	Positiva	0,000000002	0,000000001
Du et al (2023)	2023	China	Correlacional	2219	0,036	0,021	Positiva	0,090026276	0,045013138
Hou et al (2023)a	2023	China	Transversal	212	0,340	0,061	Positiva	0,000000002	0,000000001
Hou et al (2023)b	2023	China	Transversal	104	0,470	0,077	Positiva	0,000000002	0,000000001

→

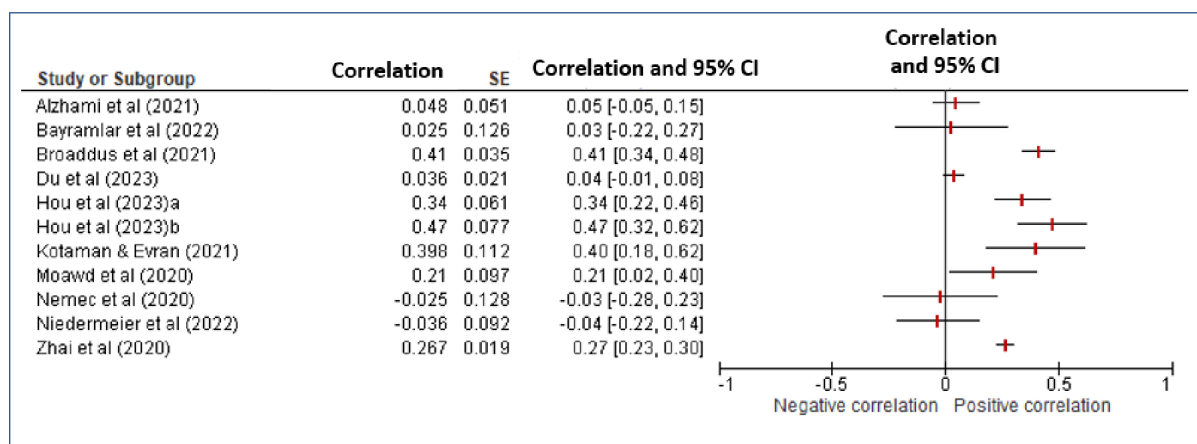
Estudio	Año	País	Diseño	N	r	ET	Dirección	p(Bilateral)	p(Unilateral)
Kotaman & Evran (2021)	2021	Turquía	Cuasi-exptal	58	0,398	0,112	Positiva	0,000813576	0,000406788
Moawd et al (2020)	2020	Arabia Saudí	Transversal	100	0,210	0,097	Positiva	0,032113128	0,016056564
Nemec et al (2020)	2020	EEUU	Transversal	63	-0,025	0,128	Negativa	0,845746104	0,577126948
Niedermeier et al (2022)	2022	EEUU	Correlacional	121	-0,036	0,092	Negativa	0,694858178	0,652570911
Zhai et al (2020)	2020	China	Transversal	2324	0,267	0,019	Positiva	0,000000002	0,000000001

$N$  = tamaño muestral.  $r$  = coeficiente de correlación entre ejercicio físico y rendimiento académico; valores positivos indican una correlación directa entre ambas variables, y viceversa.  $ET$  = error típico de  $r$ , que se calcula mediante:  $ET = (1-r^2)/\sqrt{N-2}$ .  $p(\text{Bilateral})$  = valor de probabilidad resultante de aplicar la prueba  $t$  de significación de  $r$ , asumiendo un contraste bilateral (éste es el valor que rutinariamente reportan los programas estadísticos).  $p(\text{Unilateral})$  = conversión del valor de probabilidad bilateral,  $p(\text{Bilateral})$ , en valor de probabilidad unilateral; los valores de  $p(\text{Unilateral})$  se obtienen mediante  $p(\text{Unilateral}) = p(\text{Bilateral})/2$ , para las correlaciones positivas; para las correlaciones negativas:  $p(\text{Unilateral}) = 1 - p(\text{Bilateral})/2$ . Por ejemplo, para  $r = 0,048$  (signo positivo):  $p(\text{Unilateral}) = 0,350827698/2 = 0,175413849$ ; para  $r = -0,025$  (signo negativo):  $p(\text{Unilateral}) = 1 - 0,845746104/2 = 0,577126948$ .

### 3.7. Métodos de síntesis no meta-analítica

Ya se ha planteado anteriormente cuáles pueden ser las razones por las que no debería hacerse un MA y, en su defecto, llevar a cabo una RS. No obstante, conviene recordar que las dos características que diferencian un MA de una RS tienen que ver con la medida del resultado y los métodos de síntesis. Aunque en una RS sin MA se pueden utilizar tamaños del efecto como medida del resultado (e.g., diferencias medias tipificadas, correlaciones, etc.), si no se aplican métodos estadísticos para la síntesis meta-analítica de dichos tamaños del efecto, entonces tenemos una RS y no un MA. Por tanto, una RS que utiliza tamaños del efecto como medida del resultado se caracteriza porque no calcula el efecto medio con su intervalo de confianza, ni calcula los estadísticos de heterogeneidad (estadístico  $Q$ , índice  $I^2$ , intervalo de predicción), ni comprueba el sesgo de publicación, ni aplica análisis de moderadores mediante los métodos de síntesis meta-analítica. Si se hicieran todos estos análisis, entonces la RS sería realmente un MA.

Cuando en una RS se han obtenido tamaños del efecto, una forma de presentar estos resultados es mediante la construcción de un *forest plot*. La diferencia entre un *forest plot* de una RS y el de un MA está en que en una RS el *forest plot* no reporta el tamaño del efecto medio con su intervalo de confianza, ni aporta estadísticos de heterogeneidad. En una RS el *forest plot* presenta el tamaño del efecto para cada estudio y su intervalo de confianza, tanto de forma numérica como gráfica. Para que en una RS sea posible construir un *forest plot*, es necesario disponer de cada estudio el valor del tamaño del efecto y de su error típico (o de su varianza muestral), para lo cual suele ser necesario el tamaño muestral, al menos (Ioannidis et al., 2008). Como ejemplo, valga el *forest plot* presentado en la Figura 2, construido con los datos de la Tabla 1 relativos a la RS de Rosales-Ricardo y Cáceres-Manzano (2024) sobre la relación entre ejercicio físico y rendimiento académico en universitarios. Puede observarse cómo las correlaciones exhiben una gran heterogeneidad y cómo 5 de los 11 estudios obtuvieron correlaciones que no alcanzaron la significación estadística (al incluir sus intervalos de confianza el valor 0).

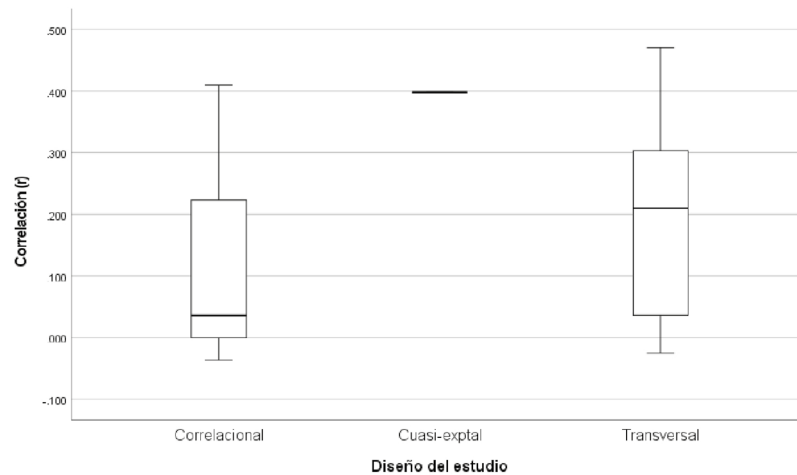


**Figura 2.** Forest plot que representa los tamaños del efecto (coeficientes de correlación,  $r$  de los estudios de la RS de Rosales-Ricardo y Cáceres-Manzano (2024) sobre la relación entre ejercicio físico y rendimiento académico en estudiantes universitarios. Elaboración propia a partir de los datos reportados en dicho artículo y la consulta de algunos de los estudios del mismo, utilizando el programa Review Manager 5.4.

Nota. SE = error típico de  $r$  (SE = 'Standard Error') 95% CI = intervalo de confianza al 95% en torno a cada correlación.

Así mismo, se puede describir la distribución de los tamaños del efecto mediante algún gráfico, como un histograma, un gráfico de caja o un diagrama en tallo y hojas, y reportar los valores mínimo, máximo, cuartiles y rango inter-cuartil exhibido por los tamaños del efecto. Obsérvese cómo éstas son herramientas estadísticas descriptivas, pero no se corresponden con los métodos de síntesis estadística propios de los MA, en los que se tiene en cuenta la variabilidad muestral de los tamaños del efecto. En la RS de Rosales-Ricardo y Cáceres-Manzano (2024), se podrían reportar las correlaciones mínima y máxima (-0,036 y 0,47, respectivamente), los cuartiles 1 y 3 (0,025 y 0,398, respectivamente) y la mediana (0,21), así como un histograma o gráfico similar.

Cuando en una RS se dispone de los tamaños del efecto, pero no se ha podido calcular los errores típicos de éstos, entonces no es posible construir un *forest plot*. En su defecto, se pueden presentar los tamaños del efecto mediante la construcción de una tabla como la presentada en la Tabla 1 de la RS de Rosales-Ricardo y Cáceres-Manzano (2024). Dicha tabla debe contener también las principales características de los estudios (e.g., año, país, tamaño muestral, tipo de diseño, población de referencia, etc.). También se pueden reportar los tamaños del efecto mínimo, máximo, percentiles, mediana, etc., como ya se ha mencionado. Alternativamente, se puede construir un diagrama de cajas múltiples (*multiple boxplot*) con objeto de comprobar la relación entre alguna característica de los estudios y los tamaños del efecto (e.g., en función de la calidad metodológica del estudio, el país, el tipo de diseño, etc.). La Figura 3 muestra un gráfico de este tipo construido con los datos de la RS de Rosales-Ricardo y Cáceres-Manzano (2024), tomando como variable moderadora el tipo de diseño (correlacional, transversal y cuasi-experimental). Puede observarse cómo la correlación mediana para los diseños correlacionales se sitúa ligeramente por encima del valor 0, mientras que los diseños transversales presentan una correlación mediana en torno a 0,20 y el único estudio cuasi-experimental se sitúa en torno al valor 0,40. Estos resultados serían indicativos de que en esta RS la correlación entre ejercicio físico y rendimiento académico puede estar condicionada por el tipo de diseño del estudio. De forma complementaria, también se puede presentar en una tabla los tamaños del efecto mínimo, máximo y la mediana para cada categoría de la variable moderadora.



**Figura 3.** Gráfico de cajas múltiples tomando el diseño del estudio como moderador y los coeficientes de correlación entre ejercicio físico y rendimiento académico como tamaños del efecto de los estudios de la RS de Rosales-Ricardo y Cáceres-Manzano (2024). Elaboración propia a partir de los datos reportados en dicho artículo y la consulta de algunos de los estudios del mismo. Se utilizó el programa IBM SPSS 28.

Una de las principales razones de llevar a cabo una RS en lugar de un MA es que no se haya podido obtener tamaños del efecto de los estudios. Como ya se comentó en el epígrafe 3.6, una alternativa es utilizar como resultado de cada estudio el valor de probabilidad unilateral obtenido en el contraste de hipótesis de interés. A partir de dichos valores de probabilidad,  $p$ , se puede aplicar el método de Fisher de acumulación de valores  $p$  mediante (cf., e.g., Borenstein et al., 2021; McKenzie y Brennan, 2019; Sánchez-Meca y Ato, 1989, p. 643, ec. 4.37):

$$\chi_{2k}^2 = -2 \sum_{i=1}^k \ln(p_i)$$

siendo  $p_i$  cada uno de los valores de probabilidad unilaterales,  $\ln$  el logaritmo natural,  $k$  el número de estudios (o de valores  $p_i$ ) y  $\chi^2$  el estadístico Chi-cuadrado de Pearson con  $2k$  grados de libertad. Este estadístico permite resolver un contraste de hipótesis en el que la hipótesis nula plantea que ninguno de los estudios presenta un efecto a favor de la hipótesis de interés, mientras que la hipótesis alternativa plantea que al menos un estudio presenta un efecto a favor de dicha hipótesis de interés. Asumiendo un nivel de significación del 5%, el contraste de hipótesis se resuelve comparando el valor de probabilidad,  $p$ , asociado al resultado de este estadístico Chi-cuadrado,  $\chi_{2k}^2$ , con el valor 0,05: si el valor  $p$  es menor que 0,05, podemos rechazar la hipótesis nula y afirmar que al menos un estudio presentó un efecto favorable a la hipótesis de interés; si el valor  $p$  es mayor que 0,05, no podemos rechazar la hipótesis nula y se concluye que ninguno de los estudios obtuvo un efecto favorable a la hipótesis de interés. Además del método de Fisher de acumulación de valores de probabilidad, existen otros tales como el método Stouffer.

Para ilustrar cómo se aplica el método de Fisher de acumulación de valores de probabilidad, la Tabla 1 recoge los valores  $p$  unilaterales de los 11 estudios de la RS de Rosales-Ricardo y Cáceres-Manzano (2024) sobre la relación entre ejercicio físico y rendimiento académico (véase la columna etiquetada ' $p$ (Unilateral)' en la Tabla 1). Debe aclararse que estos valores  $p$  unilaterales se han obtenido asumiendo que la hipótesis de partida, es decir, la que querían demostrar los autores de esta RS, era que existe una correlación positiva entre estas dos variables, de forma que aquellos valores  $p$  unilaterales bajos serían indicativos de que existe tal correlación positiva, mientras que los valores  $p$  grandes indicarían lo contrario. Es por ello que, por ejemplo, las dos correlaciones de signo negativo presentan valores  $p$  unilaterales altos (muy por encima del nivel de significación 0,05). A partir de los valores  $p$  unilaterales, y teniendo en cuenta que  $k = 11$ , aplicamos el estadístico de Fisher:

$$\chi^2_{22} = -2 \sum \ln(\pi_i) = -2 [\ln(0,175413849) + \ln(0,421624937) + \dots + \ln(0,000000001)] = -2 \cdot (-101,51) = 203,02$$

Es decir, obtenemos  $\chi^2_{22} = 203,02$ . Consultando en las tablas de la distribución Chi-cuadrado de Pearson con 22 grados de libertad, obtenemos que el valor de probabilidad asociado a dicho estadístico es muy bajo:  $p < 0,00001$ . Por tanto, podemos rechazar la hipótesis nula y concluir que al menos un estudio obtuvo un efecto a favor de la hipótesis de interés, es decir, a favor de la hipótesis de que existe una correlación positiva entre ejercicio físico y rendimiento académico.

Además de la acumulación de los valores de probabilidad, en una RS que no aporta los tamaños del efecto se puede utilizar como resultado de cada estudio 'la dirección del efecto observado en el estudio'. Si, por ejemplo, el objetivo de una RS es comprobar si las niñas de una determinada franja de edad exhiben una mayor habilidad lingüística que los niños de esa misma edad, los estudios en los que se obtenga una media más alta en dicha variable para la muestra de niñas que para la de niños diremos que la dirección del efecto es a favor de la hipótesis de interés (i.e., que las niñas muestran mayor habilidad lingüística que los niños). Los estudios en los que ocurra lo contrario, la dirección del efecto sería en contra de dicha hipótesis. Se trata, pues, de dicotomizar el resultado de cada estudio en función de la dirección del efecto: positiva (a favor de la hipótesis de interés) vs. negativa (en contra de dicha hipótesis). También puede utilizarse cuando los tamaños del efecto de los estudios están en métricas diferentes, de forma que no es posible transformarlos a una métrica común. Es muy importante tener en cuenta que la categorización de la dirección del efecto no se basa en el valor  $p$ , es decir, en si se obtuvo un resultado estadísticamente significativo (e.g.,  $p < 0,05$ ), sino en el hallazgo de un resultado favorable o desfavorable a la hipótesis de interés. Siguiendo con el ejemplo de las habilidades lingüísticas en niños y niñas, la dicotomización entre resultado favorable y desfavorable a la hipótesis de interés debe hacerse en función de las medias obtenidas por las dos muestras de niñas y niños. Si en un estudio la media en habilidades lingüísticas de las niñas fue mayor que la de los niños, entonces el efecto es favorable a la hipótesis de interés. No nos fijamos en si esa diferencia fue estadísticamente significativa (i.e.,  $p < 0,05$ ), sino en el perfil de las medias.

Una vez que tenemos registrada la dirección del efecto de cada estudio, es posible aplicar la técnica de síntesis denominada 'técnica del recuento' ('vote counting'). Esta técnica es muy simple: se obtiene el porcentaje de estudios cuya dirección del efecto es favorable a la hipótesis de interés, frente al porcentaje contrario (Boon y Thomson, 2021; Thomson y Thomas, 2013). Una forma de comparar estadísticamente estos dos porcentajes es mediante la aplicación de la prueba de los signos (o también denominada 'prueba binomial'). Esta prueba permite resolver un contraste de hipótesis en el que la hipótesis nula plantea que la proporción de estudios con resultados favorables ( $\pi_+$ ) y desfavorables ( $\pi_-$ ) a la hipótesis de interés es la misma ( $H_0: \pi_+ = \pi_-$ ), frente a la hipótesis alternativa de que la proporción de estudios con resultados favorables a la hipótesis de interés es significativamente mayor que la de resultados en contra ( $H_0: \pi_+ > \pi_-$ ). Este contraste se resuelve mediante el estadístico  $Z$  (Borenstein et al., 2021; McKenzie y Brennan, 2021; Pardo et al., 2015, pp. 242-249):

$$Z = \frac{k_+ \pm 0,5 - \frac{k}{2}}{\sqrt{\frac{k}{4}}}$$

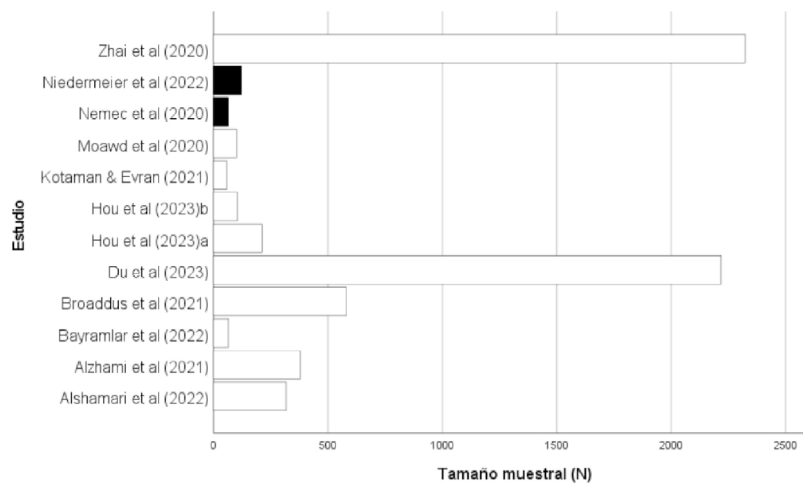
siendo  $k$  el número total de estudios y  $k_+$  el número de estudios con resultado favorable a la hipótesis de interés. Cuando  $k_+$  es mayor que  $k/2$ , a  $k_+$  se le resta 0,5, es decir,  $k_+ - 0,5$ . Cuando  $k_+$  es menor que  $k/2$ , se le suma 0,5, es decir,  $k_+ + 0,5$ . El factor  $\pm 0,5$  sólo se debe aplicar cuando  $k$  es bajo (e.g.,  $k \leq 20$ ). Bajo la hipótesis nula, el estadístico  $Z$  sigue una distribución normal tipificada. El contraste de hipótesis se resuelve comprobando si el valor de probabilidad asociado al resultado del estadístico  $Z$  es menor o mayor que el nivel de significación inicialmente planteado (e.g.,  $p < 0,05$ ): si el valor  $p$  es menor que 0,05, podemos rechazar la hipótesis nula y concluimos que el porcentaje de estudios con resultado favorable a la hipótesis de interés es significativamente superior al de estudios con resultado desfavorable. Si el valor  $p$  es mayor que 0,05, entonces no podemos rechazar la hipótesis nula, lo cual nos llevaría a concluir que el porcentaje de estudios con resultados favorables y desfavorables a la hipótesis de interés es similar y, por tanto, no habría evidencia clara de que los estudios confirmen dicha hipótesis.

En la RS de Rosales-Ricardo y Cáceres-Manzano (2024) la hipótesis de interés era que existe una correlación positiva entre ejercicio físico y rendimiento académico en universitarios. En consecuencia, las correlaciones positivas irían a favor de dicha hipótesis, mientras que las negativas irían en contra. La Tabla 1 presenta la dirección del resultado como 'positiva' o 'negativa' en la columna etiquetada 'Dirección'. Puede comprobarse cómo, de los 11 estudios ( $k = 11$ ), 9 de ellos presentaron una dirección del efecto 'positiva', es decir, a favor de la hipótesis de interés ( $k_+ = 9$ ), mientras que sólo 2 presentaron un resultado contrario a la hipótesis. Aplicando la prueba de los signos para resolver el contraste de hipótesis obtenemos:

$$Z = \frac{k \pm 0.5 - \frac{k}{2}}{\sqrt{\frac{k}{4}}} = \frac{9 - 0.5 - \frac{11}{2}}{\sqrt{\frac{11}{4}}} = \frac{3}{1.658} = 1,81$$

Consultando en las tablas de la distribución normal (o bien, aplicando la prueba de los signos con un programa estadístico al uso), obtenemos que para un valor  $Z = 1,81$ , el nivel de probabilidad (unilateral) es  $p = 0,0351$ . Al ser este valor menor que  $0,05$ , podemos rechazar la hipótesis nula de igualdad de porcentajes a favor y en contra de la hipótesis, y concluimos que esta RS presenta un porcentaje de estudios con resultado a favor de la hipótesis de interés que es significativamente mayor que el de estudios con resultados contrarios a la misma.

En una RS que utilice como resultado de cada estudio la dirección del efecto es posible utilizar un gráfico denominado 'harvest plot' (Ogilvie et al., 2008). Este gráfico consiste en presentar la dirección del efecto de cada estudio individual poniéndolo en relación con alguna característica (cuantitativa) de los estudios, tales como el tamaño muestral, la calidad metodológica, el año del estudio, etc. Se trata, pues, de un gráfico que permite analizar el posible influjo de moderadores sobre los resultados de los estudios. Como ejemplo para ilustrar este tipo de gráfico, la Figura 4 presenta un *harvest plot* con los datos de los 11 estudios de la RS de Rosales-Ricardo y Cáceres-Manzano (2024) sobre la relación entre ejercicio físico y rendimiento académico, tomando como característica de los estudios el tamaño muestral. En el gráfico se aprecia cómo los dos únicos estudios con una correlación negativa entre estas dos variables (representados con barras de color negro) se obtuvieron en sendos estudios con tamaños muestrales muy bajos.



**Figura 4.** Ejemplo de *harvest plot* tomando los datos de la RS de Rosales-Ricardo y Cáceres-Manzano (2024) sobre la relación entre ejercicio físico y rendimiento académico. Elaboración propia a partir de los datos reportados en dicho artículo y la consulta de algunos de los estudios del mismo. Se utilizó el programa IBM SPSS 28.

Nota. El eje vertical presenta los 11 estudios. El eje horizontal presenta el tamaño muestral de cada estudio. Las barras de color blanco hacen referencia a estudios cuya correlación entre las dos variables fue de signo positivo, mientras que las dos barras de color negro indican los dos únicos estudios que obtuvieron correlaciones negativas.

En resumen, los métodos de síntesis aquí presentados son los indicados en el contexto de una RS. Son métodos alternativos a los métodos de síntesis propios de los MA. Sin embargo, puede observarse que, tanto unos como otros, son procedimientos estadísticos. ¿Dónde está la diferencia entre ellos? La diferencia está en que en un MA los métodos de síntesis estadística requieren obligatoriamente utilizar tamaños del efecto, todos ellos en la misma métrica, e involucrar a la varianza de error de muestreo de dichos tamaños del efecto en el factor de ponderación, con objeto de tomar en cuenta su precisión en la estimación de los parámetros de interés. En una RS los métodos de síntesis, aun siendo también estadísticos, no pueden (o no deben) utilizar la varianza de error de muestreo de los tamaños del efecto, o bien no se dispone de los tamaños del efecto y hay que recurrir a otros indicadores estadísticos para cuantificar el resultado de cada estudio (e.g., niveles de probabilidad, dirección del efecto).

Los métodos de síntesis cuantitativa aquí presentados para las RS son procedimientos aceptados por la comunidad científica. Otros métodos menos aceptados son los que se basan en la síntesis narrativa, es decir, en utilizar como resultado de cada estudio frases que describen la evidencia sin incluir alguno de los métodos de síntesis aceptados. Estos métodos narrativos son la opción menos aconsejable porque, como ya se ha comentado antes, son muy propensos a sufrir sesgos de interpretación por parte de los revisores. Por tanto, el uso aislado de síntesis narrativas en una RS debería evitarse (Campbell et al., 2019; Popay et al., 2006).

### 3.8. Redacción de una RS

La última fase en la realización de una RS es la redacción del informe. Los apartados en los que se organiza la redacción de una RS obedecen al sistema IMRyD: Introducción, Método, Resultados y Discusión. El correcto reporte de una RS es un requisito imprescindible para que los lectores puedan valorar la confianza que pueden depositar en las conclusiones de la revisión. Campbell et al. (2019) realizaron una meta-revisión de RS realizadas en la Colaboración Cochrane y llegaron a la conclusión de que el 95% de ellas no describía con suficiente precisión los métodos de síntesis utilizados. Más recientemente, Cumpston et al. (2023) realizaron un estudio similar y también alcanzaron conclusiones análogas a las de Campbell et al. (2019). Es por ello que Campbell et al. (2020) elaboraron la guía SWiM (Synthesis Without Meta-analysis). SWiM es un checklist que consta de nueve ítems, siete de ellos para la sección Método, uno para los Resultados y uno para la Discusión. En realidad, el checklist SWiM debe aplicarse como complemento al checklist PRISMA 2020 (Page et al., 2021). PRISMA 2020 se ideó para orientar a los revisores en la correcta redacción de MA sobre la eficacia de intervenciones. Por tanto, el checklist SWiM debe utilizarse para especificar con mayor detalle cómo se definió el resultado de cada estudio y qué métodos de síntesis no meta-analíticos se aplicaron sobre éstos. En el checklist PRISMA 2020 hay ítems que sólo son aplicables a un MA, no a una RS. Por tanto, lo recomendable para redactar una RS es utilizar los dos checklists, PRISMA 2020 y SWiM, de forma combinada. En la Tabla 2 se reproduce una versión en castellano de los nueve ítems que componen el checklist SWiM.

**Tabla 2**

Ítems del checklist SWiM

<b>Método</b>	
<b>1) Agrupamiento de los estudios para la síntesis</b>	1a) Proporcione una descripción y un razonamiento de cómo se desagregaron los estudios en grupos para la síntesis (e.g., desagregación según las poblaciones, las intervenciones, los resultados, el diseño de los estudios, etc.) 1b) Detalle cualquier cambio realizado respecto del protocolo sobre cómo se desagregaron los estudios para la síntesis
<b>2) Descripción de los métodos de estandarización y transformación de los resultados</b>	Describa la/s métrica/s utilizada/s para estandarizar cada resultado. Justifique la elección de esa/s métrica/s y describa los métodos utilizados para transformar los efectos de la intervención, tal como se reportaron en los estudios, a la métrica estandarizada, citando cualquier guía metodológica utilizada
<b>3) Descripción de los métodos de síntesis</b>	Describa y justifique los métodos utilizados para sintetizar los efectos de cada resultado cuando no fue posible llevar a cabo un meta-análisis de las estimaciones de los efectos
<b>4) Criterios utilizados para priorizar los resultados para el resumen y la síntesis</b>	Donde sea aplicable, proporcione los criterios utilizados, con la conveniente justificación, para seleccionar los estudios concretos, o un estudio en particular, para la síntesis principal o para alcanzar conclusiones de la síntesis (e.g., en función del diseño de los estudios, de la evaluación del riesgo de sesgo, de la dirección del resultado respecto de la pregunta de la revisión)
<b>5) Investigación de la heterogeneidad de los efectos reportados</b>	Describa el/los método/s utilizado/s para examinar la heterogeneidad de los efectos reportados cuando no fue posible llevar a cabo un meta-análisis de las estimaciones de los efectos y sus extensiones para investigar la heterogeneidad
<b>6) Certidumbre de la evidencia</b>	Describa los métodos utilizados para valorar la certidumbre de los resultados de la síntesis
<b>7) Presentación de los datos</b>	Describa los métodos gráficos y tabulares utilizados para presentar los efectos (e.g., tablas, forest plots, harvest plots) Especifique características clave de los estudios (e.g., diseño del estudio, riesgo de sesgo) utilizadas para clasificar los estudios en el texto y en las tablas y/o gráficos, referenciando con claridad los estudios incluidos
<b>Resultados</b>	
<b>8) Presentación de resultados</b>	Para cada comparación y resultado, proporcione una descripción de los hallazgos de la síntesis y de la certidumbre de éstos. Describa los resultados en un lenguaje consistente con la pregunta objeto de interés de la síntesis y especifique qué estudios contribuyen a la síntesis
<b>Discusión</b>	
<b>9) Limitaciones de la síntesis</b>	Describa las limitaciones de los métodos de síntesis utilizados, de las formas de desagregación de los estudios utilizadas en la síntesis y cómo éstas afectan a las conclusiones que puedan extraerse respecto de la pregunta de interés de la síntesis

*Nota.* Cuando la información no se aporta en el texto de la RS, proporcione detalles acerca de dónde figura esa información (e.g., en el protocolo, en otros trabajos publicados -aporte las referencias de éstos-, o en algún sitio web -aporte la URL).

La redacción de una RS presenta varias diferencias importantes respecto de la de un MA. En primer lugar, en la sección Método de una RS debe definirse con claridad el resultado de cada estudio, que no necesariamente serán tamaños del efecto (e.g., dirección del efecto, valores  $p$ ). En segundo lugar, en la sección Método debe especificarse cuáles fueron los métodos de síntesis de los resultados, dependiendo de que los resultados se hayan definido en términos de tamaños del efecto, dirección del efecto o valores  $p$ . Estos métodos de síntesis son diferentes de los de un MA. En tercer lugar, en la sección de Resultados de una RS debe incluirse una tabla con las características principales de cada estudio (e.g., año, país, diseño, calidad metodológica, tamaño muestral, características de las muestras, etc.), así como los resultados (tamaños del efecto, dirección del efecto, valores  $p$ ). Dependiendo del tipo de resultados sintetizados, en la sección de Resultados se deben incluir e interpretar los análisis estadísticos aplicados (simples medias o medianas de los tamaños del efecto, resultado de la prueba de Fisher de acumulación de valores  $p$ , prueba de los signos), así como gráficos que ilustren la distribución de los resultados (*forest plots*, histogramas de barras, gráficos de caja, diagramas en tallo y hojas) y la relación entre éstos y características relevantes de los estudios mediante tablas desagregadas y gráficos (gráficos de cajas múltiples, *harvest plots*). Por último, las conclusiones que puede alcanzar una RS no son tan contundentes como las de un MA porque no se han aplicado métodos de síntesis estadística óptimos sobre los tamaños del efecto de los estudios. Las conclusiones de una RS tienen que ser necesariamente menos sólidas porque responden a preguntas diferentes a las de los MA. No obstante, bajo determinadas circunstancias una RS es el mejor modo de sintetizar las evidencias de un conjunto de estudios sobre una determinada pregunta.

#### 4. Conclusión

Cuando se piensa en una RS, generalmente se asume que la síntesis de los resultados de los estudios se tiene que hacer de forma meramente narrativa. Sin embargo, en este tutorial ha quedado claro que en una RS se puede hacer mucho más que simplemente ofrecer una conjetura narrativa de las evidencias. Una RS puede, y debe, aplicar métodos estadísticos a la síntesis de los resultados de los estudios. Estos métodos de síntesis, alternativos a los de los MA, permiten extraer la máxima información de los estudios y alcanzar conclusiones más precisas que la mera síntesis narrativa.

No obstante, también hay que tener en cuenta que las conclusiones que permiten alcanzar las RS son diferentes de las de los MA. Una RS que no utilice tamaños del efecto, sino la dirección del efecto (favorable vs. desfavorable a la hipótesis de interés) no permite responder a la pregunta típica de un MA: “¿Cuán grande fue la magnitud del efecto medio y cuál fue su incertidumbre y su consistencia?”. En su lugar, esa RS puede responder a una pregunta más genérica: “¿Existe evidencia clara de que hay un efecto en la población?”. Una RS que utilice como resultado de los estudios los valores de probabilidad relativos a una hipótesis de interés, tampoco permite ofrecer una conjetura razonable sobre la magnitud de los efectos, sino que ofrece una respuesta a una pregunta menos concisa: “¿Es posible afirmar que al menos un estudio de esta RS exhibió un efecto significativo a favor de la hipótesis de interés?”. Y una RS que utilice tamaños del efecto como resultado de cada estudio, incluso aunque se haya podido obtener el mismo índice del tamaño del efecto en todos ellos (e.g., correlaciones), tampoco permite ofrecer una conjetura sólida acerca de la magnitud del efecto, sino que ofrece respuestas a preguntas más genéricas: “¿Cómo se distribuyen los tamaños del efecto de los estudios?, ¿cuál es la forma de su distribución, su variabilidad y su tendencia central?”.

Por tanto, bien diseñadas, ejecutadas y redactadas, las RS constituyen una metodología excelente como alternativa a los MA, allá donde éstos no puedan aplicarse por alguno de los motivos ya mencionados. Cuando la pregunta de interés puede estudiarse empíricamente desde diferentes enfoques, perspectivas o metodologías, el carácter multidisciplinar o diverso de las evidencias debe llevar al revisor a utilizar la metodología de la RS en lugar del MA. Bajo estas circunstancias, las RS pueden ofrecer síntesis de los resultados que supondrán una excelente contribución original a la disciplina en cuestión.

No sólo es fundamental que la RS esté bien desarrollada en todas sus etapas; también es decisiva una buena redacción de la misma, que permita al lector valorar todos los aspectos de la misma y decidir sobre la validez de sus conclusiones. A este respecto, las recomendaciones aportadas por los checklists SWiM y PRISMA 2020 son de gran ayuda.

Por último, aunque no menos importante, cada vez más se está generalizando la publicación del protocolo de las RS y los MA, previa a la obtención de los resultados. Hacer públicos los objetivos y la metodología que se pretende seguir en una RS o un MA es una estrategia muy aconsejable que se enmarca dentro del movimiento de la Ciencia Abierta (Open Science Collaboration, 2015). Desde esta iniciativa, se pretende reducir los ‘grados de libertad del investigador’ para evitar prácticas de investigación cuestionables, tales como el ‘*p-hacking*’ (el investigador ‘retuerce’ los datos y/o los análisis estadísticos hasta que resultados no significativos se conviertan en significativos; Head et al., 2015), el ‘*HARKing*’ (*Hypothesis After Results are Known*: plantear hipótesis después de obtener los resultados; Murphy y Aguinis, 2019), o el ‘*reporting bias*’ (reportar sólo los resultados estadísticamente significativos, ocultando los no significativos; Ioannidis et al., 2017). Por tanto, es muy recomendable publicar el protocolo de la RS o el MA en sitios web tales como PROSPERO (<https://www.crd.york.ac.uk/prospero>) o en *Open Science Framework* (<https://help.osf.io/article/330-welcome-to-registrations>).

#### Financiación

MICIU/AEI /10.13039/501100011033/ y Fondos FEDER de la Unión Europea, Proyecto nº PID2022-137328NB-I00. Fundación Séneca de la Región de Murcia (Proyecto nº 22064/PI/22).

## Referencias

- Boon, M. H., y Thomson, H. (2021). The effect direction plot revisited: Application of the 2019 Cochrane Handbook guidance on alternative synthesis methods. *Research Synthesis Methods*, 12, 29-33. <https://doi.org/10.1002/jrsm.1458>
- Booth, A., Sutton, A., y Papaioannou, D. (2016). *Systematic approaches to a successful literature review* (2ª ed.). Sage.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., y Rothstein, H. R. (2021). *Introduction to meta-analysis* (2ª ed.). Wiley.
- Botella, J., y Caperos, J. M. (2019). *Metodología de Investigación en Psicología General Sanitaria*. Síntesis.
- Botella, J., y Sánchez-Meca, J. (2015). *Meta-análisis en ciencias sociales y de la salud*. Síntesis.
- Campbell, M., Katikireddi, S. V., Sowden, A., y Thomson, H. (2019). Lack of transparency in reporting narrative synthesis of quantitative data: A methodological assessment of systematic reviews. *Journal of Clinical Epidemiology*, 105, 1-9. <https://doi.org/10.1016/j.jclinepi.2018.08.019>
- Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., y Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *British Medical Journal*, 368, l6890. <http://dx.doi.org/10.1136/bmj.l6890>
- Cook, D. A., y West, C. P. (2012). Conducting systematic reviews in medical education: a stepwise approach. *Medical Education*, 46, 943-952. <https://doi.org/10.1111/j.1365-2923.2012.04328.x>
- Cooper, H. (2016). *Research synthesis: A step-by-step approach* (5ª ed.). Sage.
- Cooper, H., Hedges, L. V., y Valentine, J. F. (Eds.). (2019). *The handbook of research synthesis and meta-analysis* (3ª Ed.). Russell Sage Foundation.
- Cumpston, M. S., Brennan, S. E., Ryan, R., y McKenzie, J. E. (2023). Synthesis methods other than meta-analysis were commonly used but seldom specified: survey of systematic reviews. *Journal of Clinical Epidemiology*, 156, 42-52. <https://doi.org/10.1016/j.jclinepi.2023.02.003>
- Garrity, C., Nussbaumer-Streit, B., Hamel, C., Devane, D., y Cochrane Rapid Reviews Methods Group (2025). Rapid reviews methods series: assessing the appropriateness of conducting a rapid review. *BMJ Evidence-Based Medicine*, 30(1), 55-60. <https://doi.org/10.1136/bmjebm-2023-112722>
- Grant, M. J., y Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, 26, 91-108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Head, M. L., Holman, L., Lanfear, R., Kahn, A.T., y Jennions, M. D. (2015). The extent and consequences of P-Hacking in Science. *PLoS Biology*, 13(3), e1002106. <https://doi.org/10.1371/journal.pbio.1002106>
- Higgins, J. P. T., Savović, J., Page, M. J., Elbers, R. G., y Sterne, J. A. C. (2024). Chapter 8: Assessing risk of bias in a randomized trial. En J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, y V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (Vers. 6.5). Cochrane Collaboration. <https://doi.org/10.1002/9781119536604.ch8>
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., y Welch, V. A. (Eds) (2024). *Cochrane handbook for systematic reviews of interventions* (vers. 6.5). The Cochrane Collaboration. <https://doi.org/10.1002/9781119536604>
- Ioannidis, J. P., Caplan, A. L., y Dal-Ré, R. (2017). Outcome reporting bias in clinical trials: why monitoring matters. *British Medical Journal*, 356, j408. <https://doi.org/10.1136/bmj.j408>
- Ioannidis, J. P., Patsopoulos, N., y Rothstein, H. (2008). Reasons or excuses for avoiding meta-analysis in forest plots. *British Medical Journal*, 336(7658), 1413-1415. <https://doi.org/10.1136/bmj.a117>
- López-López, J. A., Rubio-Aparicio, M., y Sánchez-Meca, J. (2022). Overview of reviews: Concept and development. *Psicothema*, 34, 175-181. <https://doi.org/10.7334/psicothema2021.586>
- Lubbe, W., ten Ham-Baloyi, W., y Smit, K. (2020). The integrative literature review as a research method: A demonstration review of research on neurodevelopmental supportive care in preterm infants. *Journal of Neonatal Nursing*, 26, 308-315. <https://doi.org/10.1016/j.jnn.2020.04.006>
- McKenzie, J. E., y Brennan, S. E. (2021). Synthesizing and presenting findings using other methods. En J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, y V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (2ª ed., pp. 321-347). Wiley.
- Munn, Z., Pollock, D., Khalil, H., Alexander, L., McInerney, P., Godfrey, C. M., Peters, M., y Tricco, A. C. (2022). What are scoping reviews? Providing a formal definition of scoping reviews as a type of evidence synthesis. *JBIM Evidence Synthesis*, 20(4), 950-952. <https://doi.org/10.11124/BIIES-21-00483>
- Murphy, K. R., y Aguinis, H. (2019). HARKing: How badly can cherry-picking and question trolling produce bias in published results? *Journal of Business and Psychology*, 34, 1-17. <https://doi.org/10.1371/journal.pbio.1002106>
- Ogilvie, D., Fayter, D., Petticrew, M., Sowden, A., Thomas, S., Whitehead, M., y Worthy, G. (2008). The harvest plot: a method for synthesising evidence about the differential effects of interventions. *BMC Medical Research Methodology*, 8(8). <https://doi.org/10.1186/1471-2288-8-8>
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>

- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting. *British Medical Journal*, 372, n71. <http://dx.doi.org/10.1136/bmj.n160>
- Pardo, A., Ruiz, M. A., y San Martín, R. (2015). *Análisis de datos en ciencias sociales y de la salud I* (2ª ed.). Síntesis.
- Plakkal, N. (2023). How to write a systematic review without meta-analysis. *International Journal of Advanced Medical and Health Research*, 10, 119-23. [https://doi.org/10.4103/ijamr.ijamr\\_296\\_23](https://doi.org/10.4103/ijamr.ijamr_296_23)
- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., y Britten, N. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews*. ESRC Methods Programme.
- Rosales-Ricardo, Y., y Cáceres-Manzano, V. (2024). Effects of physical exercise on academic performance in university students: A systematic review. *Health Professions Education*, 10, 174-184. <https://doi.org/10.55890/2452-3011.1174>
- Sánchez-Meca, J. (2010). Cómo realizar una revisión sistemática y un meta-análisis. *Aula Abierta*, 38(2), 53-64.
- Sánchez-Meca, J. (2022). Revisiones sistemáticas y meta-análisis en Educación: Un tutorial. *Revista Interuniversitaria de Investigación en Tecnologías Educativas*, 13, 5-40. <https://doi.org/10.6018/riite.545451>
- Sánchez-Meca, J., y Ato, M. (1989). Meta-análisis: Una alternativa metodológica a las revisiones tradicionales de la investigación. En J. Arnau, y H. Carpintero (Eds.), *Tratado de psicología general I: Historia, teoría y método* (pp. 617-669). Alhambra.
- Sánchez-Meca, J., Marín-Martínez, F., y López-López, J. A. (2011). Meta-análisis e intervención psicosocial basada en la evidencia. *Psychosocial Intervention*, 20, 95-107. <https://doi.org/10.5093/in2011v20n1a9>
- Siddaway, A. P., Wood, A. M., y Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, 70, 747-770. <https://doi.org/10.1146/annurev-psych-010418-102803>
- Sterne, J. A. C., Hernán, M. A., McAleenan, A., Reeves, B. C., y Higgins, J. P. T. (2024). Chapter 25: Assessing risk of bias in a non-randomized study. En J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, y V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (Vers. 6.5). Cochrane. [www.cochrane.org/handbook](http://www.cochrane.org/handbook)
- Thomson, H. J., y Thomas, S. (2013). The effect direction plot: visual display of non-standardised effects across multiple outcome domains. *Research Synthesis Methods*, 4, 95-101. <https://doi.org/10.1002/jrsm.1060>
- Wells, G. A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., y Tugwell, P. (2000). *The Newcastle-Ottawa Scale (NOS) for assessing the quality if nonrandomized studies in meta-analyses*. <https://ohri.ca/en/who-we-are/core-facilities-and-platforms/ottawa-methods-centre/newcastle-ottawa-scale>