

Ejecución de revisiones sistemáticas y metaanálisis

RESUMEN

Al igual que en cualquier estudio de investigación, las revisiones sistemáticas tienen una fase conceptual, una metodológica y una de presentación de resultados. En la primera, se precisan las consideraciones que son transversales a todos los proyectos, como el planteamiento del problema, la justificación y los objetivos. Entre las consideraciones metodológicas existen muchas indicaciones, sin embargo, en este texto se argumenta que la ejecución de una revisión sistemática con o sin metaanálisis requiere de las etapas de identificación, tamización, elección e inclusión. En la primera etapa se presenta la importancia de la selección de las palabras clave, la combinación con operadores booleanos y la selección de las fuentes de información; en la segunda etapa se describe el procedimiento para aplicar criterios de inclusión conceptuales y operativos; en la tercera se presentan la evaluación de la calidad metodológica de las investigaciones y la aplicación de los criterios de exclusión; y en la cuarta y última etapa se determina el número de artículos a incluir en la revisión. Al finalizar las etapas descritas en las líneas previas, se procede con el análisis de la información y en él, se exponen los estadísticos para verificar la reproducibilidad, y hacer análisis univariado, análisis bivariado y metaanálisis.

Palabras clave: guía PRISMA, preguntas PICO, criterios de inclusión, criterios de exclusión, calidad metodológica.

¿Cómo citar este capítulo? / How to cite this chapter?

Cardona-Arias, J. A., Higueta-Gutiérrez, L. F. y Ríos Osorio, L. A. (2016). Ejecución de revisiones sistemáticas y metaanálisis. En *Revisiones sistemáticas de la literatura científica: la investigación teórica como principio para el desarrollo de la ciencia básica y aplicada* (pp. 25-40). Bogotá: Ediciones Universidad Cooperativa de Colombia. doi: <http://dx.doi.org/10.16925/9789587600377>



Performance of systematic reviews and meta-analysis

ABSTRACT

As in any research study, systematic reviews have a conceptual phase, a methodological phase, and a result presentation phase. In the first one, considerations that are cross-sectional to all projects, such as problem statement, rationale and objectives, are specified. Among methodological considerations, there are many indications; however, this text argues that the performance of a systematic review with or without meta-analysis requires the stages of identification, screening, selection, and inclusion. On the first stage, the importance of selecting keywords, the combination of Boolean operators, and the selection of sources of information are explained; on the second stage, the procedure for applying conceptual and operating inclusion criteria is described; on the third stage, the assessment of methodological research quality and the application of exclusion criteria are discussed; and, on the fourth and last stage, the number of articles to be included in the review is determined. At the end of the stages described above, the analysis of information is performed, showing the statistics to verify reproducibility and conduct univariate analysis, bivariate analysis, and meta-analysis.

Keywords: PRISMA guide, PICO questions, inclusion criteria, exclusion criteria, methodological quality.

Al igual que en cualquier estudio de investigación, las revisiones sistemáticas tienen una fase conceptual, una metodológica y una de divulgación de resultados. En la primera, se deben incluir las consideraciones que son transversales a todos los proyectos, como problema, justificación y objetivos; sin embargo, en este punto es preciso hacer explícito cuál o cuáles son las razones para no poder desarrollar la pregunta de investigación desde un estudio original, es decir, se debe precisar una de las siguientes situaciones que evidencian la particularidad de la formulación de un problema que debe abordarse desde un estudio secundario:

- * Discrepancias en estudios individuales que evalúan los mismos factores de exposición, desenlaces, intervención o prueba diagnóstica.
- * Cuando los estudios individuales no son concluyentes por presentar un bajo tamaño de muestra o resultados con baja potencia estadística.
- * Ante estudios individuales que presentan una baja validez.
- * Necesidad de evaluar si los estudios disponibles se podrían extrapolar a otras poblaciones, evitando la ejecución de investigaciones que podrían presentar conclusiones similares a las de la revisión.
- * Necesidad de tener estimaciones menos sesgadas.

En la formulación del problema de una revisión sistemática se debe tener presente que la mayoría de autores han recomendado seguir los componentes de una pregunta PICO: *population* (población), *intervention* (intervención), *comparison* (comparación), *outcome* (desenlace o resultado). En este sentido, la formulación de un problema de investigación (pregunta) debería definir de forma explícita la población de estudio, la variable respuesta o resultado, la intervención o tratamiento bajo estudio, el grupo o tipo de comparación; además, algunos autores sugieren la necesidad de indicar el tipo de estudios que se emplearán como base de la revisión sistemática (Pértega y Fernández, 2005; Guallar, Damián y Martín-Moreno, 1997).

Se debe precisar que dicha metodología es válida para estudios experimentales, en los que es explícita la intervención; sin embargo, para otro tipo de estudios, como los observacionales o los de pruebas diagnósticas, la estrategia PICO debe modificarse y en lugar de intervención se debería indicar la adaptación definida:

- * *Población*: podría ser modificado por el sujeto de la investigación que hace parte de la pregunta formulada: un agente parasitario, una prueba diagnóstica, un fenómeno de salud, entre otros.
- * *Intervención*: podría ser modificado por alguna característica relacionada con el fenómeno a evaluar, como un factor de riesgo de un fenómeno de salud; características moleculares de un agente microbiano; aspectos técnicos o ventajas y desventajas de una prueba diagnóstica, mecanismo de acción de un agente biorremediador, entre otros.
- * *Comparación*: podrían adecuarse a dicho concepto las características particulares que permiten comparar, por ejemplo, agentes microbianos, como la eficiencia en la remoción de un metal pesado; la identidad de un marcador molecular en la identificación de un agente microbiano por cierta prueba diagnóstica; la eficacia de ciertos microorganismos en la remoción de un compuesto orgánico presente en una matriz ambiental; las diferencias en las manifestaciones clínicas obtenidas en respuesta a un tratamiento farmacológico particular; los niveles de resistencia presentados por un grupo de microorganismos frente a un fármaco, entre otros.
- * *Resultado*: podrían asumirse como resultados de este tipo de adecuaciones las evidencias científicas que se derivan de la comparación propuesta en el ítem anterior, en relación con los valores de eficiencia de un método, los marcadores moleculares obtenidos frente a la necesidad de conocer los más empleados en un diagnóstico, los mecanismos de acción reportados de cierto microorganismo involucrado en un evento clínico, industrial o ambiental.

PREGUNTA DE INVESTIGACIÓN

Es importante entender que la formulación de la pregunta de investigación debe estar condicionada por la necesidad de información en un área de conocimiento específica y por la utilidad de la revisión sistemática para construir un marco teórico de referencia que dé cuenta de la información científicamente publicada y científicamente recolectada. También debe recoger las diferentes posturas teóricas que existen sobre un fenómeno del que no tenemos suficientes evidencias, no sabemos qué se ha investigado sobre él o cuáles son las tendencias que se vienen desarrollando en los últimos años al respecto, quiénes lo vienen investigando y de qué manera lo vienen haciendo. Así, la revisión sistemática se configura como una opción de sintetizar la información científica de base para el tratamiento de preguntas de investigación básica y aplicada, en un área de conocimiento específica.

OBJETIVOS DE INVESTIGACIÓN

En la ciencia, los objetivos de una investigación se constituyen en el cómo se propone resolver la pregunta de investigación y, por ende, suponen los logros a los que se espera llegar en el cumplimiento de dicha respuesta.

Existen diferentes tipos de objetivos de investigación que pueden ser formulados en la ciencia, sin embargo, en las revisiones sistemáticas, por sus características particulares, es necesario que se formulen de manera que permitan documentar de forma adecuada el fenómeno que se pretende conocer; por ello, verbos en infinitivo del tipo “describir” y “comparar” son los más empleados para proponer una forma de responder la pregunta de investigación. Recordemos que la base de las revisiones sistemáticas supone la falta de

integración de diferentes posturas teóricas respecto a una situación de interés, por tanto la descripción será la base de las respuestas a este tipo de preguntas.

ASUNTOS METODOLÓGICOS

Entre las consideraciones metodológicas existen muchas indicaciones, en las que la mayoría de textos presentan los siguientes ítems: formulación de la pregunta, búsqueda de la literatura, inclusión con flujograma de selección, evaluación de la calidad de los estudios, extracción y análisis de la información (Pértega y Fernández, 2005; Guallar et al., 1997).

Dadas las pretensiones de este libro, el cual parte de la base de que la literatura disponible presenta limitaciones en las consideraciones operativas de las revisiones, se desagregarán los componentes de la metodología de las revisiones sistemáticas en tres etapas:

1. Inclusión de los estudios en coherencia con la guía PRISMA (*Preferred reporting items for systematic reviews and meta-analyses*).
2. Análisis de la información.
3. Divulgación o presentación de resultados: componentes de un artículo de investigación derivado de una revisión sistemática.

Inclusión de los estudios: guía PRISMA

En 1999 se publicó una de las primeras guías para normalizar la presentación de las revisiones sistemáticas, particularmente las que incluían metaanálisis de ensayos clínicos controlados, conocida como la declaración QUOROM (*Quality of reporting of meta-analysis*), la cual consta de 18 puntos útiles para la evaluación de este tipo de estudios por parte de los autores y editores de revistas (Moher et al., 1999).

Esta iniciativa no logró los resultados esperados y pocos estudios demostraron su impacto en la calidad de los metaanálisis. Por ello, surgió la necesidad de actualizarla y ampliarla con nuevos ítems, por lo que en el 2009 se publicó la declaración PRISMA en la que, a diferencia de QUOROM, se tomaron en cuenta cuatro aspectos conceptuales novedosos que llevaron a la adición de nuevos ítems: i) el carácter iterativo del proceso de desarrollo de una revisión sistemática; ii) la conducción y la publicación de un estudio de investigación son conceptos distintos; iii) la evaluación del riesgo de sesgo al nivel de los estudios o de los resultados; y iv) la importancia de los sesgos relacionados con la publicación (Urrutia y Bonfill, 2010; Moher et al., 2007; Delaney et al., 2005).

En la guía PRISMA se plantean 27 puntos relacionados con la escritura y publicación de artículos de investigación derivados de revisiones sistemáticas, a lo que se suma un flujograma que orientó la selección de los estudios que se tomaron como unidades de análisis (Moher et al., 2009).

De estas dos ayudas (la lista de chequeo y el flujograma de PRISMA) se infiere que la ejecución de una revisión sistemática con o sin metaanálisis requiere de cuatro etapas:

Identificación

En esta fase se debe definir la estrategia de búsqueda de los artículos, lo cual incluye la delimitación de los términos de búsqueda (que pueden estar en uno o varios idiomas)

y las fuentes de información que se utilizarán, como bases de datos bibliográficas o buscadores abiertos. En este punto algunas revistas sugieren la presentación de todas las posibles combinaciones de términos y los operadores booleanos empleados en tal combinación. Se debe tener presente que uno de los objetivos de esta fase es que los lectores puedan disponer de una cifra de artículos publicados en relación con la pregunta de investigación, es decir, la fase de identificación es una forma de obtener una cifra que se asemeje a lo que en un estudio original sería el tamaño de la población de estudio (cuán alta o baja es la cantidad de artículos publicados en relación con el tema de investigación).

Esta etapa es de gran relevancia para definir la sensibilidad de la revisión de la literatura, la cual está implícitamente dada por la pertinencia de los términos de búsqueda, sus combinaciones y el número de fuentes consultadas (expertos en el tema de investigación deben definir los términos o se debe recurrir a bases de datos de descriptores, con términos científicamente aceptados). Cabe aclarar que en algunos temas no es posible hacer una validación de la estrategia de búsqueda por parte de los expertos o del equipo de investigación, debido a razones como que la investigación en el área sea muy novedosa.

En relación con las fuentes, es recomendable emplear mínimo una base de datos bibliográfica multidisciplinaria, una específica del área de estudio y un buscador abierto como Google Scholar. También se debe diferenciar entre buscadores, motores de búsqueda y bases de datos; así, por ejemplo, Medline es una base de datos que incluye dos buscadores, Pubmed y ovid, ScienceDirect es un motor de búsqueda y Springer Link es una base de datos.

No obstante la recomendación de emplear mínimo dos fuentes, se debe aclarar que un factor determinante de la exhaustividad es el número de bases de datos empleadas, por lo que se recomienda emplear tantas como sea posible. A esto debe sumarse las consultas en instituciones especializadas, bibliotecas y hemerotecas, contactar a investigadores que sean referentes en el tema, y otro tipo de estrategias que garanticen que los autores de la revisión consultaron todas las fuentes de información que podrían ayudar a resolver su pregunta de investigación.

Finalmente, se debe hacer explícito si la búsqueda de información se desarrolló por especificidad o sensibilidad. La primera restringida a términos mesh (Medical Subject Heading), decs (Descriptores en Ciencias de la Salud) u otros tesauros; en la segunda, las palabras clave no están restringidas a un tesauro e incluyen literatura gris o un motor de búsqueda como Google Académico.

El resultado de todas las estrategias de búsqueda (términos y fuentes) debería exportarse a una fuente común, por ejemplo un administrador de referencias (EndNote, EndNoteWeb, Zotero) que permita la eliminación de los artículos duplicados o hallados en varias fuentes o con varias estrategias de búsqueda.

Tamización

En esta fase se deben aplicar los criterios de inclusión a los estudios identificados. Su objetivo es delimitar o lograr una reducción de los estudios obtenidos en la primera fase, para lo cual se pueden establecer criterios de inclusión conceptuales o afines al tema

de investigación y criterios de inclusión relacionados con consideraciones operativas. En el primer grupo se puede tomar la delimitación temporal, características de la población, el tipo de estudio que se incluirá, las formas de medir la exposición o las variables de respuesta, los idiomas de las publicaciones, el tipo de tratamiento o exposición, el desenlace primario y cuando aplique, los desenlaces secundarios, entre otros (Pértega y Fernández, 2005; Guallar et al., 1997).

En el segundo grupo (criterios operativos), se pueden emplear los filtros propios de las fuentes consultadas. Por ejemplo, en bases de datos bibliográficas es posible hacer filtros que hacen las veces de criterios de inclusión como los siguientes:

- * *Términos de búsqueda*: que aparezcan en título, resumen o palabras clave, con esto se logra eliminar estudios que tienen los términos de búsqueda en otras partes del manuscrito sin que sean artículos afines a la pregunta de investigación. De esta forma, se da mayor especificidad a la búsqueda al limitar la selección de artículos a aquellos que den prioridad en el uso del lenguaje científico a términos más relevantes relacionados con la pregunta de investigación formulada y a las secciones más relevantes del artículo, en las cuales el uso de ciertos términos hace parte de la intención científica más especializada.
- * *Rango de tiempo*: la temporalidad de los estudios publicados, reduciéndolos a un rango de tiempo exacto, puede definirse en función del tema de investigación como consecuencia de evaluaciones preliminares de la literatura que evidencian que dichos temas están publicados acorde con un auge particular en la época científica, circunscribiéndolos a ciertos periodos, lo que nos indicaría que es en ese foco de tiempo que se debe realizar la búsqueda más sensible y específica. Entre tanto, como asunto operativo, la delimitación de un periodo de tiempo se relaciona con convenciones en las publicaciones científicas que definen cierta antigüedad de los artículos empleados para la realización de una publicación teórica, que aunque no es criterio de las revisiones sistemáticas, sí se configura en un elemento técnico que, en ausencia de asuntos científicos, pone límites al volumen de información seleccionada. En las revistas científicas que lo definen el rango delimitado es de diez años de antigüedad, que incluso se encuentra definido en el proceso técnico de los motores y bases de datos científicos.
- * *Tipo de publicación*: en la base de las revisiones sistemáticas se ha definido la especificación del tipo de estudio de las publicaciones científicas a seleccionar como uno de los criterios fundamentales y que busca eliminar sesgos de selección, garantizando homogeneidad en la información generada en dichas publicaciones. Sin embargo, desde el punto de vista práctico, y entendiendo que la investigación científica se mueve en espectros de validez y se debe contar con el principio de oportunidad de los datos, existen otros criterios a tener en cuenta para definir el tipo de publicaciones a seleccionar para una revisión sistemática. Lo primero es tener en cuenta que más allá de la explicitación de un tipo de estudio en una publicación científica también está la necesidad de verificar si aquello que se declara en el artículo sí se corresponde con una evaluación previa de cumplimiento. En este sentido, no todos los artículos clasificados en cierto tipo de estudio reflejan esa condición al ser evaluados, llevando a la necesidad de flexibilizar la delimitación del tipo de estudio de las investigaciones

seleccionadas para la revisión sistemática. Por ello, lo que se deberá definir, más que el tipo de estudio como un asunto teórico a evaluar, es la condición del tipo de información que se espera se haya producido en la investigación como base para su elección.

Con estos criterios se deben controlar varios sesgos potenciales de las revisiones sistemáticas: el de publicación, el de idioma y el de selección (Pértega y Fernández, 2005; Guallar et al., 1997).

Dado que muchas bases de datos contienen filtros que podrían hacer las veces de criterios de inclusión, esta fase se hace más ágil. En este punto se debe hacer una aclaración: si bien PRISMA recomienda la eliminación de duplicados como cierre de la fase de identificación, en algunos temas (por ejemplo, leishmaniasis, malaria, tuberculosis, diabetes etc.) el número de artículos publicado es prolijo y las búsquedas obtenidas (el número de artículos) superan el número de citas que pueden exportarse a los administradores de referencias; en este caso, para facilitar la parte operativa (y no sacrificar la exhaustividad ni la reproducibilidad de la búsqueda tratando de eliminar los duplicados en la primera fase), se podría hacer una excepción a la recomendación de PRISMA, haciendo la eliminación de duplicados toda vez que se hayan aplicado los criterios de inclusión y con ello se haya logrado una reducción del número de artículos hasta uno que sea afín a la capacidad de importación de citas desde el administrador de referencias. La otra opción es hacer una distinción en los criterios de inclusión entre operativos y académicos. Los operativos serán aplicables a la fase de identificación e incluyen la originalidad de los estudios, el idioma y el periodo de tiempo. Entre tanto, los de tipo académico se aplicarían en la lectura del *fulltext* de los artículos, verificando los elementos científicos que dan cuenta de la utilidad en la respuesta a la pregunta de investigación de los artículos tamizados.

Elección

En esta fase se deben aplicar los criterios de exclusión en los estudios que pasaron la tamización (que cumplen los criterios de inclusión). Al igual que en los estudios originales, la finalidad de la exclusión es eliminar los estudios que puedan generar sesgos en los resultados; en este sentido, se deben presentar algunos criterios relacionados con las poblaciones de estudio, la forma de medir las variables de respuesta o las de exposición, la validez interna o externa de los estudios originales, entre otros.

En la literatura disponible, es frecuente que esta etapa se denomine evaluación de la calidad metodológica de los estudios, la cual debería incluir un análisis de la confiabilidad de las mediciones y la validez de los resultados. En la validez externa se debe evaluar si el muestreo es representativo y en la validez interna se debe analizar si el tamaño de muestra es suficiente (adecuado o correcto para las estimaciones que se presenten), si se realizó control de sesgos de selección e información, cuantificación o identificación de la confusión, además de incluir otros criterios que son específicos de cada tipo de estudio:

* *Pruebas diagnósticas*: el instrumento más empleado para evaluar la calidad metodológica de este tipo de estudios es el de las guías QUADAS (Quality Assessment of Diagnostic Accuracy Studies) (Whiting et al., 2011). Este incluye 14 puntos entre los que se destacan algunos, por ejemplo, si los pacientes son representativos de quienes

recibirán la prueba en la práctica; la descripción de los criterios de selección de pacientes; si el patrón de referencia clasifica correctamente el problema de estudio; si el tiempo entre la aplicación de la prueba de estudio y la de referencia es lo suficientemente corto; la aplicación en los pacientes de la misma prueba de referencia independientemente del resultado obtenido en la prueba evaluada; la interpretación de la prueba evaluada sin conocer los resultados de la prueba de referencia y viceversa; si la información clínica disponible en la interpretación de los resultados de las pruebas es la misma que estará disponible cuando se use la prueba en la práctica; si se informan los resultados no interpretables o no concluyentes y se explican las pérdidas y retiradas del estudio (Zamora y Abaira, 2008).

- * *Estudios observacionales*: a diferencia del caso anterior, para los estudios observacionales no son tan explícitos los criterios para evaluar la calidad metodológica. Sin embargo, existen guías que se han diseñado para favorecer el reporte de este tipo de estudios, las cuales incluyen puntos relevantes para evaluar su calidad, como es el caso de las guías STROBE (Strengthening the Reporting of Observational Studies in Epidemiology), de las cuales existe una general y otras específicas para estudios transversales, casos y controles y cohortes, con 22 puntos para cada uno de los tres tipos de estudio (Von et al., 2008; Vandembroucke et al., 2009).
- * *Ensayos clínicos controlados*: en este tipo de estudios los principales criterios de evaluación de la calidad metodológica incluyen los criterios de inclusión, la aleatorización de la intervención, el ocultamiento, los cegamientos, el seguimiento de los pacientes (porcentaje de pérdidas), el análisis de homogeneidad de los grupos y el tipo de análisis (por protocolo o intención de tratar). Estos criterios han sido condensados en guías como la de Jadad et al. (1996) o la PEDRO Scale (Physiotherapy Evidence Database, 2013).

Una vez aplicada una de las guías de evaluación de la calidad metodológica, tipo lista de chequeo, se puede hacer una puntuación para que los análisis se puedan realizar agrupando los estudios de moderada y alta calidad (se asume que los de baja calidad son excluidos) como parte del análisis de sensibilidad o como parte de una metarregresión (Pértega y Fernández, 2005).

De lo anterior se desprende que la elección de artículos para una revisión sistemática sin metaanálisis debe incluir criterios de selección que involucren la evaluación de la calidad de las investigaciones frente a la pregunta de investigación definida. Para ello, sería importante el desarrollo de un instrumento con un *score* que unifique la aplicación de los criterios de selección; sin embargo, dada la posible heterogeneidad de las investigaciones que serían objeto de indagación, este instrumento de medición debe ser amplio en criterios, pero de aplicación posible a cada estudio. Así, la evaluación propuesta por el protocolo PRISMA define algunos elementos que hacen parte de la estructura del artículo, pero no define qué valoración darle a cada uno de dichos elementos. Por ello consideramos que es allí adonde debemos mirar y generar propuestas de evaluación de dicha calidad, con la intención de hacer reproducible no solo la identificación y la tamización, sino también la elección y la inclusión.

Inclusión

En esta última fase se determina el número de artículos objeto de la revisión con los cuales es posible realizar una síntesis cualitativa (la heterogeneidad en los estudios, sus mediciones de variables dependientes e independientes, impide la agrupación o generación de medidas globales) o cuantitativa (metaanálisis).

Al finalizar estas etapas se presenta el flujograma de selección de los manuscritos, indicando el tamaño de la muestra (número de artículos) de cada fase y el número de artículos no incluidos o excluidos con una breve descripción de las razones que respaldan la selección. Esto debe realizarse de forma ciega e independiente por varios evaluadores con el fin de garantizar la reproducibilidad de la selección de los estudios (Moher et al., 2009).

En caso de no hallar reproducibilidad en esta etapa, ante la falta de acuerdo entre los investigadores sobre la aplicación de los criterios de inclusión o exclusión, se debe indicar la forma de resolver las discrepancias en la selección de los artículos, lo cual generalmente se hace por consenso o por referencia a un tercero que no participó de este proceso en el primer momento (figura 5).

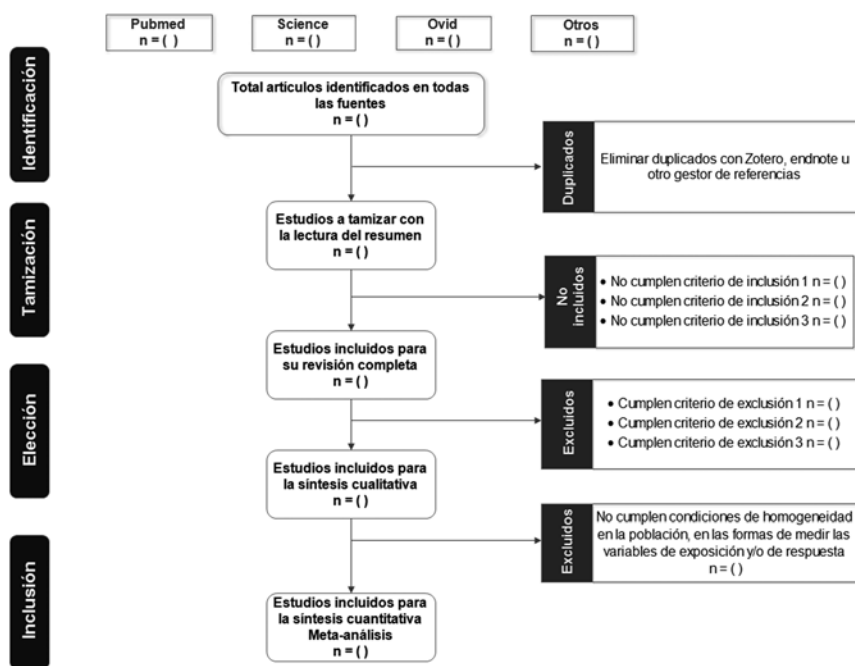


Figura 5. Algoritmo de selección de artículos. Tomada y modificada de Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement por Moher et al., 2009, *PLOS Medicine*, 6(7).

Análisis de la información

Al finalizar las fases descritas en el apartado anterior, se debe proceder con el análisis de la información. En primera instancia se deben determinar las variables y datos que se tomarán de cada estudio, el formato en que se hará la recolección de dicha información (por ejemplo, hacer una base de datos en Excel) y la forma de garantizar su reproducibilidad.



Figura 6. Análisis de la información. Elaboración propia.

En este punto es recomendable utilizar medidas estadísticas para el análisis de reproducibilidad, a diferencia de la realizada en la selección, la cual generalmente es por consenso o referencia a un tercero. Según la naturaleza de la variable analizada, se pueden calcular índices kappa para las cualitativas con el fin de establecer la concordancia de los datos extraídos por dos o más investigadores, mientras que para las variables cuantitativas se pueden calcular coeficientes de correlación intraclass, concluyendo que existe reproducibilidad en la extracción de la información de cada variable cuando estos estadísticos sean mayores a 0,95 (idealmente su resultado debería ser de 1,00).

Una vez demostrada la reproducibilidad se debe determinar el plan de análisis de la información. En este punto es pertinente precisar que, desde una perspectiva reduccionista, una revisión sistemática que presente síntesis estadística es considerada un metaanálisis. Sin embargo, tal reduccionismo resulta errado, ya que el análisis estadístico a que se hace referencia en los metaanálisis supera los estimadores y estadísticos tradicionales empleados en la parte descriptiva (uso de medidas de resumen o de frecuencias), analítica (pruebas paramétricas o no paramétricas) o explicativa (modelos de regresión). Al aludir un metaanálisis no se hace referencia a cualquier procedimiento estadístico *per se* sino a un grupo de estimadores propios de esta modalidad de investigación, los cuales se describirán después en este texto.

Recapitulando, es posible hacer buenos análisis estadísticos, principalmente descriptivos (promedios y proporciones ponderadas, análisis de tendencias, intervalos de confianza para una proporción o una media, intervalos de confianza para diferencias de promedios o proporciones, entre otros), sin que esto deje de ser una síntesis cualitativa de la revisión sistemática. Es decir, los objetivos específicos de una revisión sistemática (como describir la población de estudio, determinar la magnitud del fenómeno

de estudio, explorar o identificar algunos factores asociados) pueden implicar la necesidad de establecer algunas medidas de ocurrencia o asociación estadística, como parte de los resultados, sin que esto se corresponda con lo propio de los estimadores empleados en un metaanálisis.

Ejemplos de lo anterior se pueden encontrar en los siguientes estudios publicados por los autores de este texto:

- * *Prevalencia de la deficiencia de hierro en donantes de sangre.* Este estudio corresponde a una revisión sistemática de la literatura en la cual se combinaron estudios descriptivos en los cuales se reportaba la prevalencia de deficiencia de hierro en donantes de sangre. Con base en los resultados de los 30 estudios incluidos, se realizó el cálculo de la prevalencia global de este evento en una población de 16.979 donantes, 59,2% hombres y 40,8% mujeres; es decir, se realizó el cociente o proporción así: número de individuos con deficiencia de hierro hallados en cada estudio individual dividido la sumatoria de individuos incluidos en cada estudio y multiplicado por 100 (cálculo de una proporción de prevalencia); además, con este mismo procedimiento se calcularon las prevalencias específicas para hombres y mujeres. A las tres prevalencias (global, específica de hombres y específica de mujeres) se les calculó su respectivo intervalo de confianza del 95% para una proporción, con el estadístico $P \pm Z_{\alpha/2} * \sqrt{(PQ)/n}$, donde P corresponde a la proporción de prevalencia, $Z_{\alpha/2}$ representa el grado de confianza de la estimación (se puede tomar como una constante de 1,96 que es valor que toma una confianza del 95% en la distribución normal), Q es el complemento de P (100-P) y n es el tamaño de la muestra. En este estudio se realizaron otros análisis estadísticos y, como se dijo, esto no lo hace una revisión sistemática con metaanálisis, dado que los estadísticos empleados no son los propios de este tipo de estudio (Mantilla y Cardona-Arias, 2012).
- * *Aplicaciones de un instrumento diseñado por la OMS para la evaluación de la calidad de vida.* Se realizó una revisión sistemática con el objetivo de describir los usos del WHOQOL-BREF, una escala diseñada por la Organización Mundial de la Salud para evaluar calidad de vida. En esta investigación se incluyeron 548 estudios que cumplieron el protocolo preestablecido que analizaban una población de 234.414 sujetos. En los análisis se calcularon porcentajes con sus respectivos intervalos de confianza del 95%, de forma similar a lo citado en el estudio anterior, con el fin de hacer una buena descripción estadística (Cardona-Arias e Higuera-Gutiérrez, 2014).
- * *Prevalencia del virus papiloma humano y sus factores de riesgo en hombres.* Se realizó un revisión sistemática (Cardona-Arias, Puerta y Flórez, 2011) en la que el plan de análisis constó del cálculo de la prevalencia global de infección y las prevalencias específicas según genotipos (oncogénicos y no oncogénicos), grupo poblacional (heterosexuales, personas con VIH, militares, universitarios, homosexuales), todas con sus respectivos intervalos de confianza, igual que los dos estudios anteriores. Además, se analizaron las razones de *odds* que los estudios individuales reportaban para la infección y se realizaron intervalos de confianza para la diferencia de proporciones (con el fin de determinar diferencias estadísticas para los datos expuestos previamente) con la fórmula

$$(P1 - P2) \pm \left(Z \frac{\alpha}{2} * \sqrt{\frac{(P1 * Q1)}{n1}} + \sqrt{\frac{(P2 * Q2)}{n2}} \right)$$

En alusión a los estadísticos o estimadores propios de un metaanálisis, estos se clasifican en una tipología de cuatro fases:

Análisis de heterogeneidad

Su finalidad es determinar si los estudios individuales presentan resultados homogéneos o no; esta se mide con la Prueba de DerSimonian y Laird estadístico Q (que presenta una distribución Ji-cuadrado) y los gráficos de Galbraith y L'Abbé.

La prueba de DerSimonian y Laird se interpreta como toda prueba de hipótesis estadística, es decir, si el valor p de significación estadística es menor al error α (este es el complemento del grado de confianza, es decir, toma un valor de 0,05 cuando la confianza de las medición es del 95%), se rechaza la hipótesis nula, lo que indicaría que existe heterogeneidad, mientras que si el valor p es mayor o igual a 0,05 indicaría homogeneidad:

- * *Ho (Hipótesis nula)*: existe homogeneidad entre los estudios incluidos.
- * *Ha (Hipótesis alternativa)*: no existe homogeneidad entre los estudios incluidos.

Por su parte, la interpretación gráfica se hace mirando la distribución de los puntos en dos tipos de gráficos así:

- * *Galbraith*: si los puntos que identifican cada uno de los estudios incluidos se ubican fuera de dos líneas teóricas trazadas entre -2 y 2, indica homogeneidad. Este gráfico muestra una línea de regresión entre la precisión de cada estudio (inverso del error estándar) y el efecto estandarizado (medida de efecto dividida entre su error estándar) y dos líneas que indican el grado de confianza, con esto se determina homogeneidad. Cuando los puntos que representan cada estudio se ubican al interior de dichas líneas, los puntos por fuera de ellas evidencian que dichos estudios son los que generan mayor heterogeneidad, y como en el eje X aparece la precisión, los puntos que se encuentran hacia la derecha serán aquellos con mayor precisión (Galbraith, 1988).
- * *L'Abbé*: se determina si los puntos están cerca o lejos de una línea teórica que indica la homogeneidad. Este se emplea para los casos en que la variable de respuesta es dicotómica, dado que este gráfico muestra la frecuencia del evento en el grupo expuesto en contraste con la frecuencia de los no expuestos. Por ello, la diagonal separa los estudios así: por encima estarán aquellos que indican que el evento se presenta con mayor frecuencia en los expuestos y en la parte inferior aquellos en que el evento se presentó con mayor frecuencia en los no expuestos. Si los puntos se ubican a ambos lados, es reflejo de heterogeneidad (L'Abbé, Detsky y O'Rourke, 1987).

Se debe determinar si existe heterogeneidad y tratar de explicar su origen (confusión, sesgos), controlarla (metarregresión) y luego aplicar el modelo de análisis (Pértega

y Fernández, 2005). Si no existe heterogeneidad, el metaanálisis se realiza bajo efectos fijos, en caso contrario bajo efectos aleatorios.

Sesgo de publicación

Se evalúa con las pruebas de Egger (Estadístico t) y de Begg (Estadístico Z) y los gráficos de Egger y *funnel plot*.

- * Ho: no existe sesgo de publicación.
- * Ha: existe sesgo de publicación.

Por tanto, siempre que el valor p sea menor que el α (0,05 para una confianza del 95%) se determinará que existe sesgo de publicación.

El *funnel plot* es el complemento gráfico de la prueba de Begg. En este gráfico se relacionan el tamaño de la muestra y el tamaño del efecto detectado bajo el supuesto de que si todos los estudios detectan un efecto similar y presentan una dispersión diferente en razón del tamaño de la muestra, los puntos graficados deberían formar un “embudo” con datos superiores e inferiores a una línea teórica que se traza alrededor del cero. En caso de existir un sesgo de publicación, —según la presunción de que los estudios con resultados negativos (que no corroboran la hipótesis o presentan una baja magnitud del efecto) y los que presentan bajo tamaño de muestra no son publicados o tienen menos posibilidad de ser publicados—, se estaría frente a un gráfico en el que solo habrían hallazgos favorables o que corroboraron la hipótesis de investigación, de tal forma que en el gráfico no habría valores por encima y por debajo de cero. Dada la subjetividad de este recurso se recomienda hacer las pruebas de Begg o Egger (Begg y Mazumbar, 1994).

Análisis de sensibilidad

El análisis de sensibilidad estudia la influencia individual de cada estudio en el resultado global para determinar si los resultados están sesgados por estudios con escasa calidad metodológica, por trabajos no publicados o que no cumplan estrictamente los criterios de selección. Es replicar la revisión quitando en cada paso uno de los estudios incluidos, para ver si se obtienen o no resultados similares de forma global (Pértega y Fernández, 2005; Guallar et al., 1997).

Resultados globales

En el análisis global se puede realizar un metaanálisis acumulado, un gráfico de influencia para evaluar la contribución de cada estudio en el resultado global y en el metaanálisis final con el *forest plot*.

Si no existe heterogeneidad entre los resultados de los estudios individuales, se hace análisis de efectos fijos, que indica que el efecto global obtenido del metaanálisis es igual a los estudios individuales y se asume que la única fuente de variabilidad de los resultados es la intraestudios, atribuible al azar. Si se presenta heterogeneidad entre los estudios, se usa el análisis de efectos aleatorios que indica que el efecto de cada estudio varía;

la variación incluye diferencias en la población, en los métodos o los tratamientos, y además consideran dos posibles fuentes de variabilidad intraestudio e interestudios. Cuando existe una gran heterogeneidad no es apropiado calcular efectos globales, sino que se recomienda realizar análisis por subgrupos (Pérttega y Fernández, 2005; Guallar et al., 1997). Si en los resultados globales o agrupados se presenta asimetría, esta podría atribuirse a un sesgo de selección y de publicación, a un sesgo de localización (idioma, sesgo de citación, sesgo de múltiple publicación), a una verdadera heterogeneidad, al tamaño de muestra, al pobre diseño metodológico de los estudios, a análisis inadecuados o a asociaciones artefactuales (Cabello, 2005).

REFERENCIAS

- Begg, C. y Mazumbar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50, 1088-1101.
- Cardona-Arias, J. e Higueta-Gutiérrez, L. (2014). Impacto del VIH/SIDA sobre la calidad de vida: metaanálisis 2002-2012. *Revista Española de Salud Pública*, 88(1), 85-99.
- Cardona-Arias, J., Puerta, J. y Flórez, J. (2011). Prevalencia del virus papiloma humano y sus factores de riesgo en hombres: revisión sistemática. *Infectio*, 15(4), 268-276.
- Delaney, A., Bagshaw, S., Ferland, A., Manns, B., Laupland, K. y Doig, C. (2005). A systematic evaluation of the quality of meta-analyses in the critical care literature. *Critical Care*, 9, 575-582.
- Galbraith, R. (1988). A note on graphical presentation of estimated odds ratios from several clinical trials. *Statistics in Medicine*, 7, 889-894.
- Guallar, E., Damián, J. y Martín-Moreno, J. (1997). Metaanálisis y revisiones sistemáticas en cardiología. *Revista Española de Cardiología*, 50, 345-354.
- Jadad, A., Moore, R., Carroll, D., Jenkinson, C., Reynolds, D., Gavaghan, D. y McQuay, H. (1996). Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clinical Trials*, 17, 1-12.
- L'Abbé, K., Detsky, A. y O'Rourke, K. (1987). Meta-analysis in clinical research. *Annals of Internal Medicine*, 107, 224-233.
- Mantilla, C. y Cardona-Arias, J. (2013). Metaanálisis: prevalencia de deficiencia de hierro en donantes de sangre repetitivos y asociación con sexo, 2001-2011. *Revista Cubana Hematología, Inmunología y Hemoterapia*, 29(1), 59-72.
- Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D. F. et al. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: The QUOROM statement. *Lancet*, 354: 1896-1900.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. y The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, 6(7). doi:10.1371/journal.pmed.1000097
- Moher, D., Tetzlaff, J., Tricco, A., Sampson, M. y Altman, D. (2007). Epidemiology and reporting characteristics of systematic reviews. *PLOS Medicine*, 4(3), e78. doi:10.1371/journal.pmed.0040078
- Pérttega, S. y Fernández, S. (2005). Revisiones sistemáticas y metaanálisis. *Cad Aten Primaria*, 12(3), 166-171.

- Physiotherapy Evidence Database PEDRO. (2013). Recuperado de www.pedro.org.au.
- Rada, G., Andrade, M., Leyton, V., Pacheco, C. y Ramos, E. (2004). Medicina basada en evidencia. *Revista Médica de Chile*, 132, 253-259.
- Urrutia, G. y Bonfill, X. (2010). Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Medicina Clínica (Barc)*, 135(11), 507-511.
- Vandenbroucke, J., Von, E., Altman, D., Gøtzsche, P., Mulrow, C., Pocock, S. Poole, C., Schlesselman, J., Egger, M., y en nombre de la Iniciativa STROBE. (2009). Mejorar la comunicación de estudios observacionales en epidemiología (STROBE): explicación y elaboración. *Gaceta Sanitaria*, 23(2), 1-28.
- Von, E., Altman, D., Egger, M., Pocock, S., Gøtzsche, P., Vandenbroucke, J. y STROBE Initiative. (2008). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Journal of Clinical Epidemiology*, 61(4), 344-349.
- Whiting, P., Rutjes, A., Westwood, M., Mallett, S., Deeks, J., Reitsma, J. Leeflang, M., Sterne, J., Bossuyt, P., y QUADAS-2 Group. (2011). QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine*, 155, 529-536.
- Zamora, J. y Abaira, V. (2008). Análisis de la calidad de los estudios de evaluación de pruebas diagnósticas. *Nefrología*, 28(2), 42-45.