



Medicina de Familia SEMERGEN

www.elsevier.es/semergen



FORMACIÓN CONTINUADA - METODOLOGÍA Y TÉCNICAS

Revisiones sistemáticas en cinco pasos: III. Cómo evaluar la calidad de los estudios

K.S. Khan^{a,b}, A. Bueno Cavanillas^{a,b,*} y J. Zamora^{b,c,d}

^a Departamento de Medicina Preventiva y Salud Pública, Universidad de Granada, Granada, España

^b CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, España

^c Unidad de Bioestadística Clínica, Hospital Ramón y Cajal, Madrid, España

^d Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, Reino Unido

Recibido el 31 de diciembre de 2021; aceptado el 14 de mayo de 2022

Disponible en Internet el 5 de junio de 2022

PALABRAS CLAVE

Revisión sistemática;
Riesgo de sesgo;
Lista de
comprobación;
Evaluación de
calidad;
Heterogeneidad

Resumen La evaluación de la calidad de los estudios desempeña un papel fundamental en cada paso del proceso de realización de una revisión sistemática. Para realizar una evaluación precisa se utilizan listas de comprobación que incluyen las cuestiones genéricas relevantes al diseño de los estudios incluidos en la revisión. Los elementos de la lista utilizada pueden modificarse en relación con los componentes específicos que definen la pregunta de la revisión. La evaluación de calidad se utiliza para describir los estudios seleccionados, explicar la heterogeneidad, decidir acerca de la idoneidad de un metaanálisis, evaluar la solidez de la evidencia recopilada (existencia de sesgos) y formular recomendaciones para futuras investigaciones.

© 2022 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

KEYWORDS

Systematic review;
Risk of bias;
Checklist;
Quality assessment;
Heterogeneity

Systematic reviews in five steps: III. Assessing the quality of the literature

Abstract Study quality assessment plays a key role in every step of the process when conducting a systematic review. Checklists are used to perform an accurate quality assessment. They include generic questions relevant to the design of the studies included in the review. The checklist items can be modified in relation to the question-specific components related to participants, interventions and outcomes. Quality assessment is used to describe the selected studies, explain heterogeneity, decide on the feasibility of undertaking a meta-analysis, assess the strength of the collated evidence, and make recommendations for future research.

© 2022 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Published by Elsevier España, S.L.U. All rights reserved.

* Autor para correspondencia.

Correo electrónico: abueno@ugr.es (A. Bueno Cavanillas).

Introducción

La calidad de los estudios es el talón de Aquiles de una revisión sistemática; las conclusiones obtenidas a partir de estudios realizados con escaso rigor científico no pueden tener el mismo valor que las que se apoyan en estudios correctamente realizados^{1,2}. La calidad de un estudio integra múltiples componentes, desde la justificación hasta las conclusiones, pasando, por supuesto, por el desarrollo metodológico.

Los componentes evaluados con mayor frecuencia son las medidas empleadas para minimizar el sesgo y garantizar la adecuación del diseño del estudio, su realización y análisis. Generalmente se establece un umbral específico para definir el tipo de diseño más débil que se aceptará en el proceso de evaluación, garantizando así un nivel mínimo de calidad para todos los estudios incluidos en la revisión sistemática. Una vez seleccionados los estudios que cumplen los criterios de inclusión, incluyendo una calidad aceptable en el diseño, se analizará en profundidad la calidad de cada uno de ellos. El uso de listas de comprobación facilita la evaluación de la calidad de los estudios o, dicho de otra forma, la evaluación del riesgo de sesgo. El objetivo es evaluar las características del estudio en cuanto a su diseño, ejecución y análisis, determinar el riesgo de sesgo y obtener un parámetro que califique o cuantifique la calidad de la evidencia de una manera más precisa. En la [tabla 1](#) se incluye un breve glosario en el que se definen los términos específicos relacionados con la evaluación de calidad de los estudios incluidos en una revisión sistemática.

Listas de comprobación para evaluar la calidad de los estudios

La evaluación de la calidad metodológica se basa en la valoración de las características del diseño, su ejecución y análisis. Se utilizan listas de comprobación y existen multitud de herramientas publicadas para ello o, en definitiva, para la evaluación del riesgo de sesgo, en las que se pueden identificar los elementos anteriores. Parte de las listas de comprobación disponibles establecen una escala numérica con el objetivo de asignar una puntuación cuantitativa de la calidad de cada estudio; otras clasifican los estudios en subgrupos según su calidad sea «alta», «media» o «baja», de acuerdo con el grado en el que se cumplen los criterios de calidad. Algunas de estas listas de comprobación carecen del rigor científico necesario. Por este motivo, es preciso examinar si todos los elementos de las listas publicadas son pertinentes o si se consideran otros que, al ser relevantes para la revisión en curso, los investigadores consideran que deberían incluirse. De hecho, es probable que los valores numéricos asignados a cada elemento, o la arbitrariedad de los criterios que establecen la dicotomía «alta o baja» para evaluar la calidad, no se ajusten a las necesidades del proyecto. En algunos casos, es posible encontrar listas que no necesitan modificación alguna, puesto que evalúan con precisión la calidad de los estudios de una materia concreta. Sin duda, esto ayudará a mejorar la comparabilidad con otros estudios sobre el mismo tema. No obstante, teniendo en cuenta la complejidad de la evaluación, con

frecuencia es preferible, e incluso altamente recomendable, adaptar las listas existentes a las características de la revisión, considerando la naturaleza de la pregunta formulada al inicio.

¿Cómo se pueden reconocer los elementos adecuados para la evaluación? Hay que tener en cuenta que los estudios pertinentes son susceptibles de presentar sesgos específicos, relacionados con el modo en el que se obtuvieron y analizaron los datos. En estos casos será necesario plantearse la posibilidad de modificar una lista de comprobación genérica para incluir los elementos adicionales que se consideren apropiados y excluir aquellos que sean prescindibles. Suele resultar útil consultar las guías publicadas acerca de cómo valorar críticamente las publicaciones en materia de salud ([tabla 2](#)).

Sesgos principales contemplados en las listas de comprobación genéricas

La mayor parte de las listas de comprobación consideran varios sesgos genéricos³⁻⁵. En este subapartado se tratarán 4 tipos de sesgos que tienen un impacto sobre la validez de un estudio: sesgo de selección, sesgo de ejecución, sesgo de clasificación o sesgo de información (diagnóstico) y sesgo de deserción (pérdidas) ([fig. 1](#)).

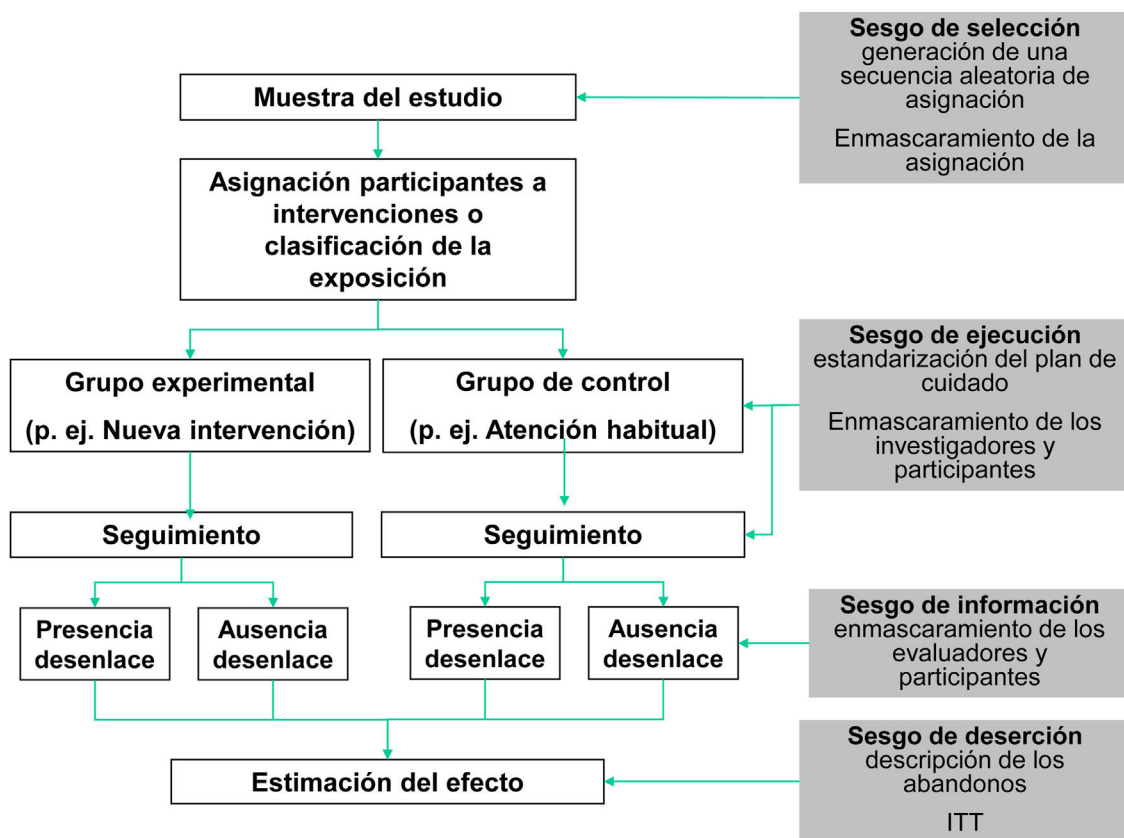
Para garantizar la validez de un estudio que mide la eficacia de una intervención, es imprescindible, en primer lugar, que los grupos de comparación sean razonablemente iguales al inicio del estudio. Si los grupos de comparación presentan una distribución desigual de las características relacionadas con el pronóstico, no se podrá saber si el efecto detectado sobre el desenlace puede atribuirse o no a la intervención, podría deberse exclusivamente a las diferencias iniciales en el pronóstico. Técnicamente, esto se debe a la presencia de sesgos de confusión introducidos por una mala selección de los participantes. El sesgo de selección surge, en los estudios observacionales, al reclutar los participantes, y en los estudios experimentales, en el momento de asignar los participantes a los grupos de estudio. La valoración de este sesgo requiere comprobar si se han tomado las medidas necesarias para minimizarlo o, si es posible, evitarlo. Al realizar una revisión sistemática, se tiende a preferir los estudios experimentales con asignación aleatoria de los participantes. La aleatorización es el mecanismo preferido para controlar los sesgos de confusión, incluso los introducidos por variables desconocidas, o no medidas, relacionadas con el pronóstico.

Una vez asignados los participantes, las intervenciones o las exposiciones que no formen parte de la investigación pueden introducir nuevos sesgos, llamados sesgos de ejecución o de cointervención. Para detectarlos, es necesario evaluar si los planes de tratamiento estaban correctamente estandarizados y si los participantes y los investigadores ignoran el grupo (experimental o control) al que están asignados.

El sesgo de clasificación está presente, sobre todo cuando los desenlaces evaluados se miden de forma subjetiva y si los participantes y los evaluadores que determinan la presencia o no del desenlace conocen el grupo de asignación. Por otra parte, si los estudios seleccionados valoran diferentes desenlaces, hay que estar atento a la prioridad de cada

Tabla 1 Glosario sobre términos del tercer paso en una revisión sistemática

Término	Definición
Heterogeneidad	El grado de diferencia en la magnitud del efecto entre estudios individuales. Para realizar un juicio clínico sobre la heterogeneidad hay que observar las diferencias entre participantes, intervenciones o exposiciones y en el desenlace de los estudios
Lista de comprobación	Relación de características que permiten verificar la correcta aplicación de la metodología en una investigación
Sesgo de selección	Diferencias sistemáticas en el pronóstico o en la sensibilidad terapéutica inicial entre grupos de estudio
Sesgo de ejecución (o de cointervención)	Diferencias sistemáticas en la atención prestada a los sujetos del estudio, aparte de las intervenciones protocolizadas
Sesgo de información (o de clasificación)	Diferencias sistemáticas entre los grupos debidas a los procedimientos de evaluación del desenlace o a su interpretación
Sesgo de deserción	Diferencias sistemáticas entre grupos de estudio debido a la exclusión del estudio o el abandono voluntario de los participantes
Análisis de intención de tratar	En inglés, <i>Intention to Treat Analysis</i> (ITT). Se trata de un análisis en que se analizan los sujetos en función del grupo inicial que les fue asignado, independientemente de si abandonaron, cumplieron con la intervención, cambiaron de grupo o recibieron intervenciones alternativas. Un ITT auténtico incluye un desenlace (observado o estimado) para todos los participantes


Figura 1 Flujo de un estudio con los sesgos principales.

uno de ellos^{6,7}. En ocasiones, tanto autores como editores se decantan por la publicación de resultados estadísticamente significativos en lugar de respetar el orden de importancia de los posibles desenlaces descrito en el protocolo. Estas desviaciones del protocolo son indicios de estudios de baja

calidad, que solo se pueden evaluar cuando el protocolo está previamente registrado. Esta es una de las razones para que tanto el registro prospectivo como la publicación del protocolo se consideren de gran utilidad para evaluar la calidad de una investigación.

Tabla 2 Ejemplo de una lista de comprobación para evaluar la calidad de un estudio**1. Definir la pregunta clínica**

Ejemplo: ¿Un tratamiento antiestrogénico aumenta las posibilidades de embarazo en parejas infértiles cuando la subfertilidad se debe al factor del hombre?

2. Definir los criterios de selección:

Naturaleza de la pregunta

Evaluación de la efectividad clínica

Diseño del estudio

Estudios que permiten comparaciones entre al menos 2 grupos

Umbral de calidad

Criterio de inclusión: estudios experimentales

Criterio de exclusión: estudios observacionales

3. Identificar una lista para comprobar la calidad de los estudios y adaptarla a las necesidades del proyecto

Elementos de calidad

Secuencia aleatoria para asignar los pacientes a las intervenciones

Adecuado

Números generados aleatoriamente por ordenador o tablas de números aleatorias

Inadecuado

Alternancia, número de registro de casos, fechas de nacimiento o días de la semana

Poco claro o no figura

Ocultación de la secuencia aleatoria

Adecuado

Aleatorización centralizada a tiempo real o controlada por un servicio externo, por ejemplo, el servicio de

Farmacia

Otros enfoques con métodos sólidos para ocultar la secuencia aleatoria a los investigadores y participantes

Inadecuado

Alternancia, número de registro de casos, fechas de nacimiento o días de la semana, listas abiertas de números aleatorios o sobres numerados en series (incluso sobres opacos sellados pueden ser objeto de manipulación)

Poco claro o no figura

Doble ciego

Adecuado

Ni los investigadores ni los participantes en el estudio conocen el grupo al que cada participante está

asignado

Inadecuado

Los investigadores o los participantes en el estudio conocen el grupo al que cada participante está asignado

Poco claro o no expresado

Descripción de los abandonos (permite un análisis de la intención de tratar)

Adecuado

Incluye en el análisis todas las personas que abandonaron o cuyo seguimiento se perdió

Se aportan datos numéricos y razones de abandono para cada grupo

La descripción permite realizar un análisis con base en el principio de ITT

Inadecuado

Solo se dan números de pérdidas (no razones) en cada grupo

La descripción no permite realizar un análisis con base en el principio del ITT

Poco claro o no figura

4) Incorporar la evaluación de calidad en la revisión

Para informar sobre la calidad de los estudios incluidos en la revisión

Para ayudar a determinar la solidez de las inferencias

Las pérdidas o abandonos del estudio implican que los grupos dejan de ser comparables. Se conoce como sesgo de deserción. Recoger los datos de todos los participantes permite realizar un análisis por intención de tratar (ITT, por sus siglas en inglés), que ayuda a controlar el efecto del sesgo de deserción. Los desenlaces que presentan los participantes se analizan según el grupo al que fueron asignados, sin tener en cuenta si siguieron o no la intervención hasta el final, cambiaron de grupo o abandonaron el estudio antes de

completarlo. Si los estudios seleccionados no analizan sus datos de esta forma, es posible realizar un análisis complementario siempre que los autores proporcionen una descripción de los abandonos y pérdidas con información suficiente para estimar los desenlaces que se presentaron entre ellos, cualquiera que sea el motivo por el que salieron del estudio. Sin esta información no sería posible realizar este tipo de análisis, pero siempre se puede hacer un análisis de sensibilidad mediante el que se valora como cambiarían

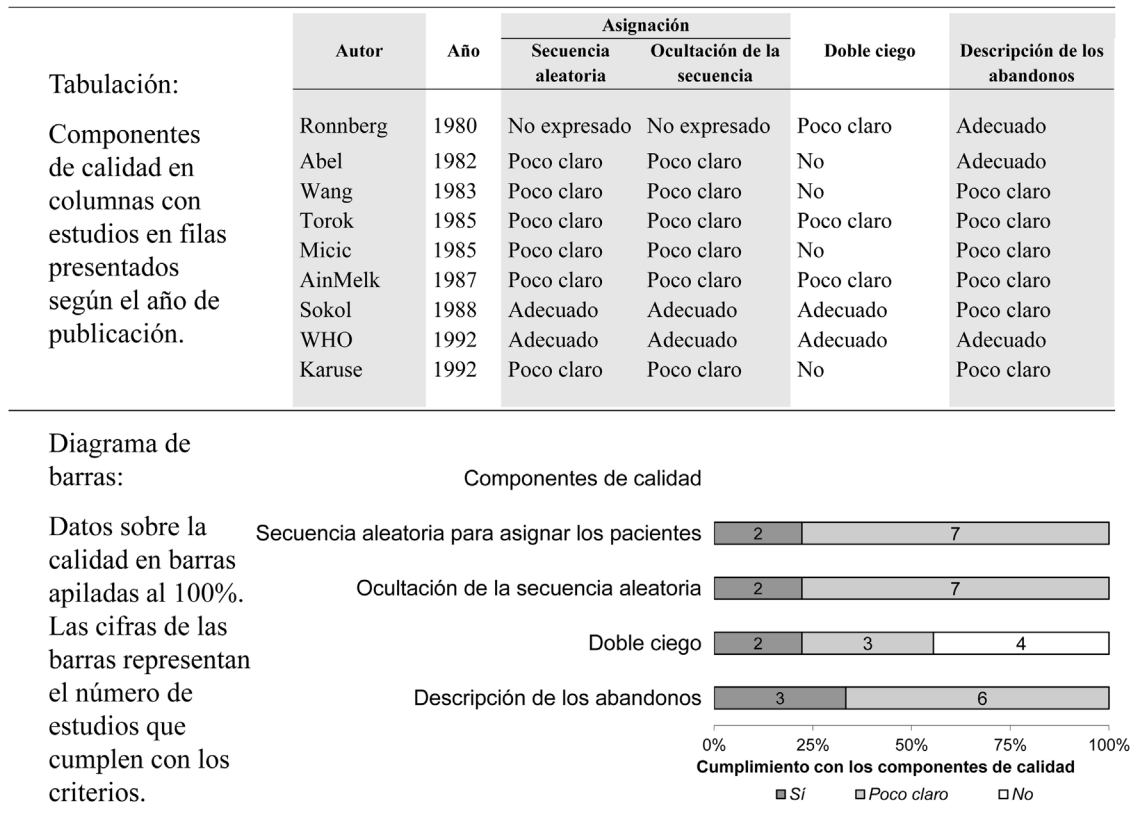


Figura 2 Tabulación y construcción de un diagrama para presentar los componentes de calidad de los estudios incluidos en una revisión sistemática.

los resultados atribuyendo el mejor o peor efecto posible a las observaciones ausentes. La [figura 2](#) contiene un ejemplo de evaluación de calidad de los estudios incluidos en una revisión sistemática⁸.

Evaluación de calidad en revisiones que incluyen estudios con diseños diferentes

En el pasado se consideraron aceptables exclusivamente las revisiones sistemáticas de ensayos aleatorizados controlados. Sin embargo, son muchas las ocasiones en las que no se dispone de este tipo de estudios, bien porque no serían factibles o bien porque hasta la fecha no se han realizado. En estos casos puede ser muy útil la revisión sistemática de estudios observacionales, pero también se pueden incluir diferentes diseños cuando la pregunta de investigación así lo requiere. Por ejemplo, si el objetivo se centra solo en la evaluación de la eficacia de una intervención será preferible centrarse en ensayos aleatorizados y controlados, pero si al mismo tiempo se pretende evaluar los efectos adversos de dicha intervención, habrá que incluir estudios de cohortes e incluso estudios de casos y controles cuando se trata de efectos adversos poco frecuentes. La evaluación de la calidad de estos estudios requerirá diferentes listas de comprobación, adaptadas a las peculiaridades, y en particular al riesgo de sesgo específico de cada uno de los diseños incluidos.

Fiabilidad de la lista de comprobación

Dadas las diferencias en la precisión con la que se presenta la información en los diferentes estudios, el protocolo de la revisión debería describir con claridad cómo se va a evaluar la calidad. Esto implica la creación de formularios de extracción de datos con una codificación coherente de las respuestas. A ser posible, los formularios deberían pasar por una fase piloto ejecutada de forma independiente por varios revisores. Contrastando los datos recogidos por cada uno de ellos se puede estimar la fiabilidad del proceso de evaluación de la calidad en una muestra de estudios, antes de aplicar dicho formulario a la totalidad de ellos. Para que esto sea posible, lo idóneo es realizar el proyecto en colaboración con otros coautores.

Presentación de las evaluaciones de calidad en una revisión

Existen diversas formas de presentar la información acerca de cómo los estudios incluidos en la revisión se adecuan a los componentes de calidad incluidos en la lista de comprobación. La elaboración de una tabla en la que se recoja la información aportada por cada uno de los estudios es la más clara. Ordenar los estudios de acuerdo con su calidad resulta más complejo. La manera más sencilla es realizar una clasificación en función del número de componentes que cumplan. Los problemas surgen cuando los estudios cumplen

con el mismo número de componentes, pero las deficiencias se encuentran en distintas áreas. En estos casos, los estudios cuyas deficiencias sugieran un mayor potencial de sesgo deben ocupar un grado más bajo en la clasificación. Es importante tener en cuenta que no existe ningún criterio universal a la hora de clasificar los estudios, la relevancia de los distintos componentes de calidad puede diferir entre tema y tema. Por ejemplo, el enmascaramiento es esencial en estudios con desenlaces subjetivos, pero no tanto en los que se basan en desenlaces objetivos. Es por eso que, en cada revisión, los investigadores deben ser capaces de decidir cómo clasificar los estudios, según su calidad, en un contexto propio.

Conclusiones

Ante la diversidad de los estudios hallados en el segundo paso de una revisión sistemática, es vital utilizar una lista de comprobación de la calidad de los estudios adecuada a cada proyecto. Asimismo, es fundamental la identificación de los sesgos más probables en cada uno de los diseños incluidos, esto va a suponer una gran ventaja para evaluar posteriormente su calidad. Se recomienda tabular los datos extraídos en la evaluación de calidad con el objetivo de poder clasificar con rigor los estudios en función del grado de cumplimiento de los componentes de calidad definidos a priori.

Financiación

Los autores no han recibido ningún tipo de financiación para la elaboración de este documento.

Conflicto de intereses

Todos los autores declaran no tener ningún conflicto de intereses relacionado con este manuscrito.

Agradecimientos

Agradecemos a Daniel Gavilán Cabello su colaboración en la traducción de este artículo.

Khalid S. Khan está contratado por la Universidad de Granada como investigador distinguido gracias a una ayuda del programa Beatriz Galindo del Ministerio de Ciencia, Innovación y Universidades. España.

Bibliografía

1. Khan KS, Kunz R, Kleijnen J, Antes G. *Systematic Reviews to Support Evidence-Based Medicine*. 2nd ed. CRC Press; 2011.
2. Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. *J R Soc Med*. 2003;96:118–21, <http://dx.doi.org/10.1258/JRSM.96.3.118>.
3. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane handbook for systematic reviews of interventions*. 2nd ed. John Wiley & Sons; 2019.
4. Tufanaru C, Munn Z, Aromataris E, Campbell J, Hopp L. Chapter 3: Systematic reviews of effectiveness. En: Aromataris E, Munn Z, editores. *JBIM Manual for Evidence Synthesis*. 2020. p. 71–88, <http://dx.doi.org/10.46658/JBIMES-20-04>.
5. Centre for Reviews and Dissemination. Core principles and methods for conducting a systematic review of health interventions. En: *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. York: York Publishing Services Ltd. 2009. p. 1-108; [consultado 25 Dic 2021]. Disponible en: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf
6. Khan KS, Bueno-Cavanillas A, Zamora J. Revisiones sistemáticas en cinco pasos: I. Cómo formular una pregunta para la que se pueda obtener una respuesta válida. *Med Fam Semer*. 2022, <http://dx.doi.org/10.1016/J.SEMERG.2021.12.006>. Published online May 3, 2022.
7. Khan KS, Bueno-Cavanillas A, Zamora J. Revisiones sistemáticas en cinco pasos: II. Cómo identificar los estudios relevantes. *SEMERGEN*. 2022 [En prensa].
8. Khan KS, Daya S, Jadad AR. The importance of quality of primary studies in producing unbiased systematic reviews. *Arch Intern Med*. 1996;156:661–6, <http://dx.doi.org/10.1001/ARCHINTE.1996.00440060089011>.