

UNIVERSIDAD DE GRANADA
FACULTAD DE MEDICINA

**La inteligencia artificial fue poco utilizada
pero relevante en las revisiones
sistemáticas sobre el COVID-19:
*un estudio metodológico***

Artificial intelligence in COVID-19 evidence
syntheses was underutilized, but impactful:
a methodological study

Journal of Clinical Epidemiology 148 (2022) 124-134

DOI [10.1016/j.jclinepi.2022.04.027](https://doi.org/10.1016/j.jclinepi.2022.04.027)



Medicina Preventiva y Salud Pública

Trabajo Fin de Grado

13 de junio de 2022

Alumno:

D. Juan Ramón

TERCERO HIDALGO

Tutor:

Prof. D. Khalid

Saeed KHAN

Índice

Información académica	IV
Compromiso de originalidad	VI
Autorización para la defensa	VII
Abstract	IX
Artículo en español	1
1. INTRODUCCIÓN	1
2. MATERIALES Y MÉTODOS	2
2.1. Búsqueda y selección de revisiones	2
2.2. Extracción de datos	2
2.3. Síntesis de datos	3
3. RESULTADOS	4
3.1. Búsqueda y selección de revisiones	4
3.2. Descripción de las revisiones incluidas	4
3.3. Comparación de las revisiones con IA con los controles	4
3.4. Descripción narrativa de los usos de la IA	6
Asistencia en la búsqueda bibliográfica	11
Filtrado de ensayos controlados aleatorizados	11
Automatización del <i>screening</i>	11
4. DISCUSIÓN	12
Referencias	15
Apéndice A. Artículo original	19
Apéndice B. Checklist PRISMA 2020	31
Apéndice C. Palabras clave de búsqueda	37
Apéndice D. Evaluación metodológica AMSTAR-2	38
Apéndice E. Normas de publicación de la revista	39

Información académica

Título en español	<i>“La inteligencia artificial fue poco utilizada pero relevante en las revisiones sistemáticas sobre el COVID–19: un estudio metodológico”</i>
Título original	<i>“Artificial intelligence in COVID–19 evidence syntheses was underutilized, but impactful: a methodological study”</i>
Alumno	Juan Ramón Tercero Hidalgo (jrterceroh@correo.ugr.es)
Tutor	Prof. D. Khalid Saeed Khan (profkkhan@gmail.com)
Departamento	Medicina Preventiva y Salud Pública
Curso académico	2021/2022 (convocatoria ordinaria de junio)
Tipo de trabajo	Trabajo de investigación original, formato artículo
Ref. publicación	Journal of Clinical Epidemiology 148 (2022) 124-134
Factor de Impacto	6.437 (JCR 2020) 1º decil en <i>Public, Environmental & Occupational Health</i>
Identificador DOI	10.1016/j.jclinepi.2022.04.027
Fecha publicación	2 de mayo de 2022
Núm. de palabras	3785

Versión original y material suplementario

La versión original (en inglés) del presente trabajo se encuentra publicada (con DOI [10.1016/j.jclinepi.2022.04.027](https://doi.org/10.1016/j.jclinepi.2022.04.027)) en el **JOURNAL OF CLINICAL EPIDEMIOLOGY**, revista con un 2020 *Impact Factor* de 6.437 y 1º decil en *Public, Environmental & Occupational Health*. Se adjunta a este documento como **Apéndice A**.

El **material suplementario 1** se adjunta traducido como Apéndices **B, C y D**. El material suplementario 2 (base de datos en Excel) está **disponible en la web de la revista**. Las normas de publicación de la revista se adjuntan como **Apéndice E**.

Información sobre financiación y conflictos de intereses

La elaboración de este estudio ha sido becada con una “*Beca de Iniciación a la Investigación para Estudiantes de Grado*” del Plan Propio de Investigación 2021 de la UGR. El coste de la publicación *open-access* ha sido financiado por la Universidad de Granada y el Consorcio de Bibliotecas Universitarias de Andalucía (CBUA).

No hay conflictos de intereses que declarar.

Protocolo previo, disponibilidad de los datos y Comité de Ética

El protocolo de este estudio fue registrado de forma prospectiva en Open Science Forum Registries (DOI [10.17605/OSF.IO/H5DAW](https://doi.org/10.17605/OSF.IO/H5DAW)) con fecha 11 de mayo de 2021.

Los datos primarios de este estudio (junto con instrucciones para replicar su análisis) están disponibles en Dryad (DOI [10.5061/dryad.9kd51c5j6](https://doi.org/10.5061/dryad.9kd51c5j6)), y se obtuvieron de los siguientes recursos disponibles en el dominio público: *Living Overview of the Evidence, Unpaywall* y *2020 Journal Citation Reports*.

Por utilizar únicamente datos disponibles en el dominio público, la realización de este estudio no requirió de autorización previa de un Comité de Ética.

Autoría y aportaciones

Bajo supervisión del Prof. D. Khalid Saeed Khan, el alumno fue el principal responsable de la recogida de datos, análisis, y elaboración del escrito final de la presente publicación. El alumno también ejerció de *corresponding author*, y se encargó de comunicarse con las distintas revistas científicas hasta la aceptación del artículo.

En la realización del estudio han participado además los siguientes coautores:

Coautor	Filiación principal
Prof. D. Khalid Saeed Khan (<i>tutor</i>) Prof. ^a Dña. Aurora Bueno Cavanillas Prof. ^a Dña. Carmen Amezcua Prieto D. Rodrigo Fernández López	Departamento de Medicina Preventiva y Salud Pública, Universidad de Granada
Prof. D. Juan Francisco Huete Guadix Prof. D. Juan Manuel Fernández Luna	Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada
Prof. D. Javier Zamora Romero	Unidad de Bioestadística Clínica, Hospital Universitario Ramón y Cajal (IRYCIS)

Compromiso de originalidad

JUAN RAMÓN TERCERO HIDALGO, con DNI 77654524S, declara que:

El presente Trabajo Fin de Grado (TFG), con título «*Artificial intelligence in COVID-19 evidence syntheses was underutilized, but impactful: a methodological study*», es original, no habiéndose utilizado fuentes sin ser citadas debidamente ni haberse presentado previamente para su evaluación como TFG o por cualquier otro motivo. De no cumplir con este compromiso, soy consciente que, de acuerdo con la **Normativa de Evaluación de Calificación de los estudiantes de la Universidad de Granada** aprobada por Acuerdo del Consejo de Gobierno de 20 de mayo de 2013 (BOUGR núm. 71, de 27 de mayo de 2013) y modificada por los Acuerdos del Consejo de Gobierno de 3 de febrero de 2014 (BOUGR núm. 78, de 10 de febrero de 2014); de 23 de junio de 2014 (BOUGR núm.83, de 25 de junio de 2014) y de 26 de octubre de 2016 (BOUGR núm. 112, de 9 de noviembre de 2016), esto «conllevará automáticamente la calificación numérica de cero [...] independientemente del resto de las calificaciones que el estudiante hubiera obtenido. Esta consecuencia debe entenderse sin perjuicio de las responsabilidades disciplinarias en las que pudieran incurrir los estudiantes que plagien» (Art. 15. Punto 2).

Y para que conste así lo firmo el día 5 de junio de 2022.

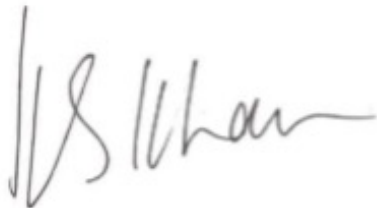
Autorización para la defensa

D. KHALID SAEED KHAN, profesor del Departamento de Medicina Preventiva y Salud Pública de la Universidad de Granada:

Como tutor del Trabajo Fin de Grado de **JUAN RAMÓN TERCERO HIDALGO**, titulado «*Artificial intelligence in COVID-19 evidence syntheses was underutilized, but impactful: a methodological study*»,

Informa de que ha sido realizado y redactado por el mencionado alumno, bajo mi tutela y dirección, y con esta fecha autorizo a su depósito y defensa pública.

En Granada, a 5 de junio de 2022

A handwritten signature in black ink, appearing to read 'Khalid Saeed Khan', written in a cursive style.

Khalid Saeed Khan

Abstract

Objective: A rapidly developing scenario like a pandemic requires the prompt production of high-quality systematic reviews, which can be automated using artificial intelligence (AI) techniques. We evaluated the application of AI tools in COVID-19 evidence syntheses.

Study design: After prospective registration of the review protocol, we automated the download of all open-access COVID-19 systematic reviews in the *COVID-19 Living Overview of Evidence* database, indexed them for AI-related keywords, and located those that used AI tools. We compared their journals' JCR Impact Factor, citations per month, screening workloads, completion times (from pre-registration to preprint or submission to a journal) and AMSTAR-2 methodology assessments (maximum score 13 points) with a set of publication date matched control reviews without AI.

Results: Of the 3 999 COVID-19 reviews, 28 (0.7%, 95% CI 0.47-1.03%) made use of AI. On average, compared to controls ($n = 64$), AI reviews were published in journals with higher Impact Factors (median 8.9 vs. 3.5, $P < 0.001$), and screened more abstracts per author (302.2 vs. 140.3, $P = 0.009$) and per included study (189.0 vs. 365.8, $P < 0.001$) while inspecting less full texts per author (5.3 vs. 14.0, $P = 0.005$). No differences were found in citation counts (0.5 vs. 0.6, $P = 0.600$), inspected full texts per included study (3.8 vs. 3.4, $P = 0.481$), completion times (74.0 vs. 123.0, $P = 0.205$) or AMSTAR-2 (7.5 vs. 6.3, $P = 0.119$).

Conclusion: AI was an underutilized tool in COVID-19 systematic reviews. Its usage, compared to reviews without AI, was associated with more efficient screening of literature and higher publication impact. There is scope for the application of AI in automating systematic reviews.

MeSH keywords:

Artificial Intelligence; Systematic Review; COVID-19; Automation; Research Design; Bibliometrics

Abstract en español

Objetivo: Un escenario dinámico como una pandemia requiere la rápida producción de revisiones sistemáticas de calidad, que pueden automatizarse utilizando inteligencia artificial (IA). Se evaluó el uso de herramientas de IA en las revisiones sistemáticas sobre COVID-19.

Diseño del estudio: Tras el registro prospectivo del protocolo del estudio, automatizamos la descarga de todas las revisiones sistemáticas *open-access* sobre COVID-19 en la base de datos *COVID-19 Living Overview of Evidence*, las indexamos en busca de palabras clave relacionadas con la IA y localizamos aquellas que utilizaban herramientas de IA. Comparamos el factor de impacto de sus revistas, las citas por mes recibidas, las cargas de trabajo en *screening*, el tiempo de elaboración (días desde el registro del protocolo hasta el primer *preprint* o envío a una revista) y la evaluación metodológica AMSTAR-2 (máximo, 13 puntos) con un grupo control de revisiones sistemáticas que no usaron IA emparejadas por fecha de publicación.

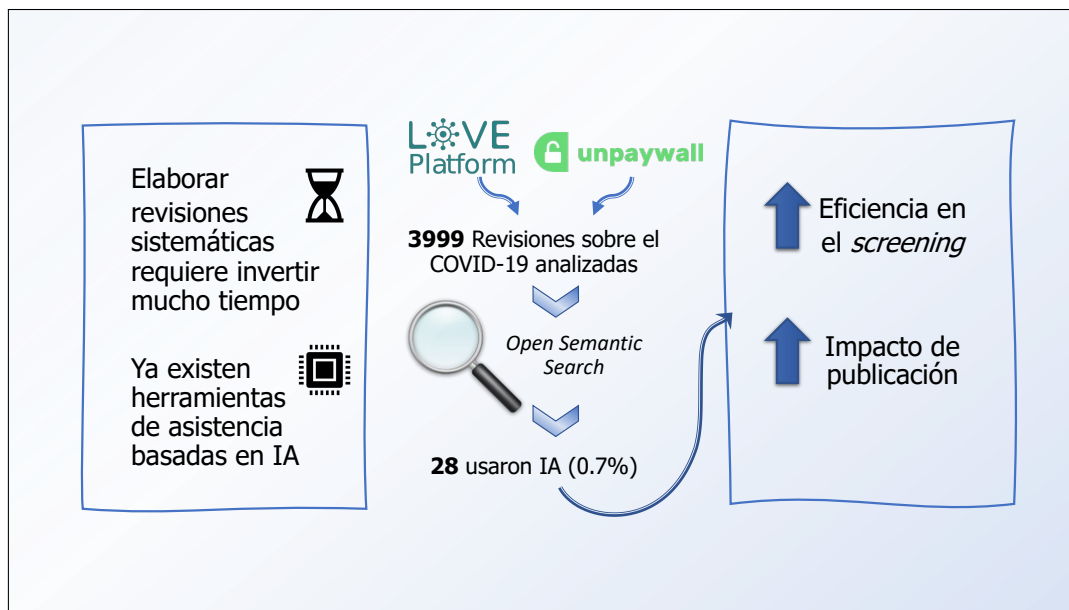
Resultados: De las 3999 revisiones sobre COVID-19, 28 (0,7%, IC al 95%: 0,47-1,03%) hicieron uso de IA. De media, en comparación con los controles ($n = 64$), las revisiones con IA se publicaron en revistas con mayor factor de impacto (mediana 8,9 vs. 3,5, $P < 0,001$), y examinaron más abstracts por autor (302,2 vs. 140,3, $P = 0,009$) y por estudio incluido (189,0 vs. 365,8, $P < 0,001$), a la vez que inspeccionaron menos *full texts* por autor (5,3 vs. 14,0, $P = 0,005$). No se encontraron diferencias en las citas recibidas (0,5 vs. 0,6, $P = 0,600$), en *full texts* inspeccionados por estudio incluido (3,8 vs. 3,4, $P = 0,481$), en los tiempos de elaboración (74 frente a 123, $P = 0,205$) ni en puntuación AMSTAR-2 (7,5 frente a 6,3, $P = 0,119$).

Conclusión: La IA fue una herramienta infrutilizada en las revisiones sistemáticas sobre COVID-19. Su uso, en comparación con las revisiones sin IA, se asoció con una selección más eficiente de la literatura y un mayor impacto de publicación. Hay cabida para la aplicación de la IA en la automatización de las revisiones sistemáticas.

Palabras clave MeSH:

Inteligencia Artificial; Revisión Sistemática; COVID-19; Automatización; Diseño de Estudios; Bibliometría

Visual abstract



¿Qué aporta este estudio?

Hallazgos clave

- El uso de la inteligencia artificial (IA) en las revisiones sistemáticas de COVID-19 fue muy bajo.
- Las revisiones de COVID-19 que utilizaron herramientas de IA mostraron un mayor impacto de publicación y una reducción en la carga de trabajo.

¿Qué añade esto a lo que se sabía?

- El cribado semiautomatizado y el filtrado de ECA son los casos más notables de uso de las herramientas de IA en la síntesis de pruebas.
- Faltan herramientas de revisión sistemática que integren la IA de forma cohesionada.

¿Cuál es la implicación y qué debería cambiar ahora?

- Hay cabida para la aplicación de la IA en la automatización de revisiones sistemáticas en el futuro.

Artículo en español

1. INTRODUCCIÓN

La medicina basada en la evidencia depende de la ágil producción de revisiones sistemáticas para orientar y actualizar la práctica clínica y las políticas sanitarias [1]. Esta es una tarea exigente y costosa, pues requiere que equipos de varios revisores consulten múltiples repositorios y bases de datos, examinen miles de citas y artículos potencialmente relevantes, extraigan los datos pertinentes de los estudios seleccionados y sinteticen sus resultados [2, 3]. En el contexto de la pandemia de SARS-CoV2/COVID-19, se necesitaban urgentemente métodos para acelerar este laborioso proceso [4, 5].

La elaboración de revisiones sistemáticas requiere seguir procedimientos consistentes y estandarizados para obtener resultados fiables. Sin embargo, la necesidad de acelerar la obtención de resultados durante la pandemia se tradujo en una disminución generalizada de la calidad metodológica de las revisiones [6, 7] y la popularización de las “revisiones rápidas” [8, 9] (que acortan los plazos habituales de producción sacrificando en rigor de la búsqueda, precisión del *screening* o la extracción de datos y a costa de un mayor riesgo de sesgos). ¿Son estas omisiones inevitables para obtener resultados más rápidos?

En contraste, las soluciones basadas en inteligencia artificial (IA) automatizan partes del flujo de trabajo imitando la resolución humana de problemas (incluyendo el uso de *machine-learning*, procesamiento del lenguaje natural, minería de datos y otros subcampos) [10] para complementar o sustituir los esfuerzos humanos con un riesgo limitado de sesgos [11, 12, 13], y se han empleado previamente pero de forma escasa [14] en revisiones sistemáticas con el fin de mejorar el *screening* [15] y la extracción de datos [16, 17]. Su objetivo es acortar los tiempos de producción, permitir un cribado más amplio de la literatura y reducir la carga de trabajo de los revisores sin comprometer la calidad metodológica.

En este trabajo, evaluamos el uso de las herramientas basadas en IA en las revisiones sistemáticas sobre el COVID-19 para determinar empíricamente si, en comparación con las revisiones sobre el COVID-19 sin IA, tuvieron un impacto en la producción, la calidad y la publicación de las revisiones sistemáticas.

2. MATERIALES Y MÉTODOS

Este estudio metodológico [18] se ha elaborado de acuerdo con las directrices PRISMA 2020 [19] (checklist proporcionada como [Apéndice B](#)), y su protocolo se registró en *Open Science Forum Registries* (DOI [10.17605/OSF.IO/H5DAW](#)) [20] de forma prospectiva.

2.1. Búsqueda y selección de revisiones

Se consideraron para su inclusión todas las revisiones sistemáticas relacionadas con el COVID-19 que pudieran haber hecho uso de cualquier herramienta de IA (*machine-learning*, *deep-learning* o procesamiento del lenguaje natural) para acelerar, mejorar o complementar cualquier aspecto de la realización de la revisión (búsqueda, *screening*, extracción de datos y síntesis). Implementamos un *script* (disponible en el DOI [10.5061/dryad.9kd51c5j6](#)) [21] para procesar todas las referencias bibliográficas registradas en la base de datos *COVID-19 Living Overview of Evidence (L-OVE)* [22] filtrando aquellas clasificadas como “revisión sistemática” entre el 1 de diciembre de 2019 y el 15 de agosto de 2021, y, a continuación, consultar la base de datos *Unpaywall* [23] con cada DOI extraído para obtener un registro JSON con enlaces de descarga. El proceso se repitió 3 veces desde la publicación de nuestro protocolo con el objetivo de reducir la potencial pérdida puntual de revisiones por errores de conexión con dichos servidores (la última búsqueda fue el 17 de agosto de 2021).

Para seleccionar las revisiones que utilizaron IA, elaboramos una lista de palabras clave con alta probabilidad de aparecer en artículos con herramientas de IA ([Apéndice C](#)). Indexamos cada archivo descargado con el motor de búsqueda *OpenSemanticSearch*, que ejecutamos en una máquina virtual local. Cada archivo que contenía alguna de nuestras palabras clave fue inspeccionado manualmente de forma independiente por dos autores (JRTH y RFL). Se incluyeron *preprints* y artículos en otros idiomas además del inglés. El único criterio de exclusión aplicado fue la no disponibilidad de los archivos en modalidad *open-access*, debido a la necesidad de evaluar la sección de métodos de cada revisión incluida. Para crear un grupo de comparación con suficiente poder estadístico de revisiones sin IA, por cada revisión incluida se utilizaron los registros obtenidos para seleccionar aleatoriamente 3 controles con la misma fecha de publicación (con un margen de un día si no había suficientes revisiones disponibles en una fecha determinada). Además, localizamos e incluimos en el análisis todas las versiones anteriores de las revisiones categorizadas como *living* o como actualizaciones de otra publicada previamente (*updated*).

2.2. Extracción de datos

Dos autores (JRTH y RFL) extrajeron manualmente los siguientes datos de cada revisión: tipo de revisión (según la descripción de sus autores: estándar, *rapid* o *scoping*, *living* o *updated*); información sobre la financiación recibida y los conflictos de

intereses declarados; estado de publicación, factor de impacto en el *Journal Citation Reports (JCR)* de 2020 de la revista que la publica y número de citas recibidas (hasta el 17 de agosto de 2021); número de abstracts inspeccionados, *full texts* revisados y estudios primarios incluidos; número de autores y de revisores que participaron en el *screening*; y fechas de registro del protocolo (si estaba disponible) y de la versión más antigua de la revisión. Para las revisiones *living* y *updated*, se calculó la diferencia de abstracts inspeccionados y estudios incluidos entre cada una de sus versiones y se atribuyó su recuento de citas a la más reciente (para evitar contabilizarlas doblemente). Se utilizó Excel para registrar todas las variables.

Tres autores (JRTH y RFL, asistidos por CAP) evaluaron todas las revisiones con la herramienta AMSTAR-2 de calidad metodológica y riesgo de sesgos [24]. Se excluyeron sus ítems 11-12 y 15, aplicables únicamente a meta-análisis (según lo preestablecido en nuestro protocolo) y se otorgaron 0,5 puntos a las respuestas “parcialmente sí” cuando correspondiera, permitiendo así una puntuación máxima de 13 puntos. Para las revisiones *living* y *updated*, sólo se evaluó su versión más reciente (para evitar contabilizarlas doblemente). En las revisiones que incluían tanto ensayos controlados aleatorizados como estudios observacionales, la pregunta 9 (evaluación del riesgo de sesgo de los estudios individuales) se evaluó por separado para cada tipo de estudio. La lista de evaluación utilizada se proporciona como [Apéndice D](#).

2.3. Síntesis de datos

Se calcularon los ratios de abstracts y *full texts* inspeccionados por autor (como medida de la carga de trabajo) y por estudio incluido (como precisión del *screening*). El número de revisores que participaron en el *screening* se reportó de forma inconsistente entre los estudios y, por tanto, no se utilizó en los cálculos. Se calculó el tiempo de elaboración de las revisiones con protocolo pre-registrado como la diferencia entre la fecha de su protocolo y la fecha de publicación del primer *preprint* (o la fecha de recepción en la revista, en el caso de los artículos publicados sin *preprint* disponible). Los tiempos de elaboración de las revisiones *living* y *updated* se calcularon como la diferencia entre las fechas de publicación de cada una de sus versiones. Se excluyeron de esta variable las revisiones no pre-registradas debido a la heterogeneidad con que se informaron sus fechas de inicio. Se utilizó la prueba de chi-cuadrado de Pearson para comparar el porcentaje de revisiones rápidas, *living*, que recibieron financiación y que fueron publicadas entre los grupos. El Factor de Impacto JCR de las revistas publicadas, los recuentos de citas recibidas, las cargas de trabajo de la revisión, los tiempos de elaboración y las calificaciones AMSTAR-2 se reportaron como medianas con rangos intercuartílicos (IQR), se representaron mediante diagramas *box-and-whisker*, y se compararon mediante el test de Wilcoxon-Mann-Whitney. Se utilizó el software R (versión 4.0.5) para el cálculo estadístico y GraphPad Prism 9.2.0 para los gráficos. También se proporcionó una descripción narrativa de las revisiones que utilizaron inteligencia artificial, detallando qué partes

del proceso de la revisión se automatizaron y qué software fue utilizado, cómo difirieron las calificaciones de AMSTAR-2 entre ellas, y cómo los autores justificaron o qué impacto atribuyeron al uso de las herramientas basadas en IA.

3. RESULTADOS

3.1. Búsqueda y selección de revisiones

Como se muestra en la *Figura 1*, identificamos 7 050 registros bibliográficos de revisiones sistemáticas sobre el COVID-19, descargamos con éxito 3 999 e inspeccionamos manualmente 580 que contenían algunas de nuestras palabras clave. Seleccionamos 20 revisiones, de las cuales localizamos 8 versiones anteriores, sumando así 28 revisiones (0,7% del total, intervalo de confianza al 95%: 0,47-1,03%) con uso de IA. De las 60 revisiones seleccionadas como controles según su fecha de publicación, localizamos otras 4 versiones anteriores, lo que hace un total de 64 revisiones sin uso de IA. La lista completa de revisiones sistemáticas seleccionadas se proporciona en formato Excel (*Supplementary Material 2* en la publicación original; sección “*Revisiones incluidas*”) con todas las variables extraídas y el desglose de la herramienta AMSTAR-2 con cada ítem evaluado. También se proporciona la lista completa de revisiones inspeccionadas manualmente y finalmente no incluidas en el trabajo (sección “*Revisiones excluidas*”).

3.2. Descripción de las revisiones incluidas

Las variables extraídas se resumen en la *Tabla 1* y pueden visualizarse en la *Figura 2*. De las 20 revisiones seleccionadas por utilizar IA, hubo 5 *rapid reviews* (25%, con 1 *scoping review* y 1 *rapid evidence map*) y 5 revisiones *living* (25%). Quince revisiones proporcionaron una declaración de conflictos de intereses, de las cuales 12 (60%) declararon haber recibido financiación externa; 12 (60%) habían sido publicadas. De las 60 revisiones del grupo control, hubo 6 *rapid reviews* (10%, con 1 *scoping review*) y 3 revisiones *living* (5%). Cincuenta y siete revisiones proporcionaron una declaración de conflicto de intereses, de las cuales 27 (45%) declararon haber recibido financiación externa; 48 (80%) habían sido publicadas. Los Factores de Impacto JCR y los recuentos de citas recibidas mostraron una alta variabilidad en el grupo IA, principalmente debido a la inclusión de 3 revisiones publicadas en BMJ [25, 26, 27], 2 revisiones Cochrane [28, 29] y 1 revisión de la revista Lancet [30]. Además, sólo 10 revisiones en el grupo IA (50%) y 22 en los controles (36%) registraron previamente un protocolo, lo que en conjunto supuso únicamente 44 registros para el cálculo de los tiempos de elaboración.

3.3. Comparación de las revisiones con IA con los controles

El grupo IA incluyó más revisiones *living* que los controles (5/20 vs. 3/60, IC del 95% diferencia absoluta 0,2 a 39,8%, $P=0,010$), pero no mostró diferencias en

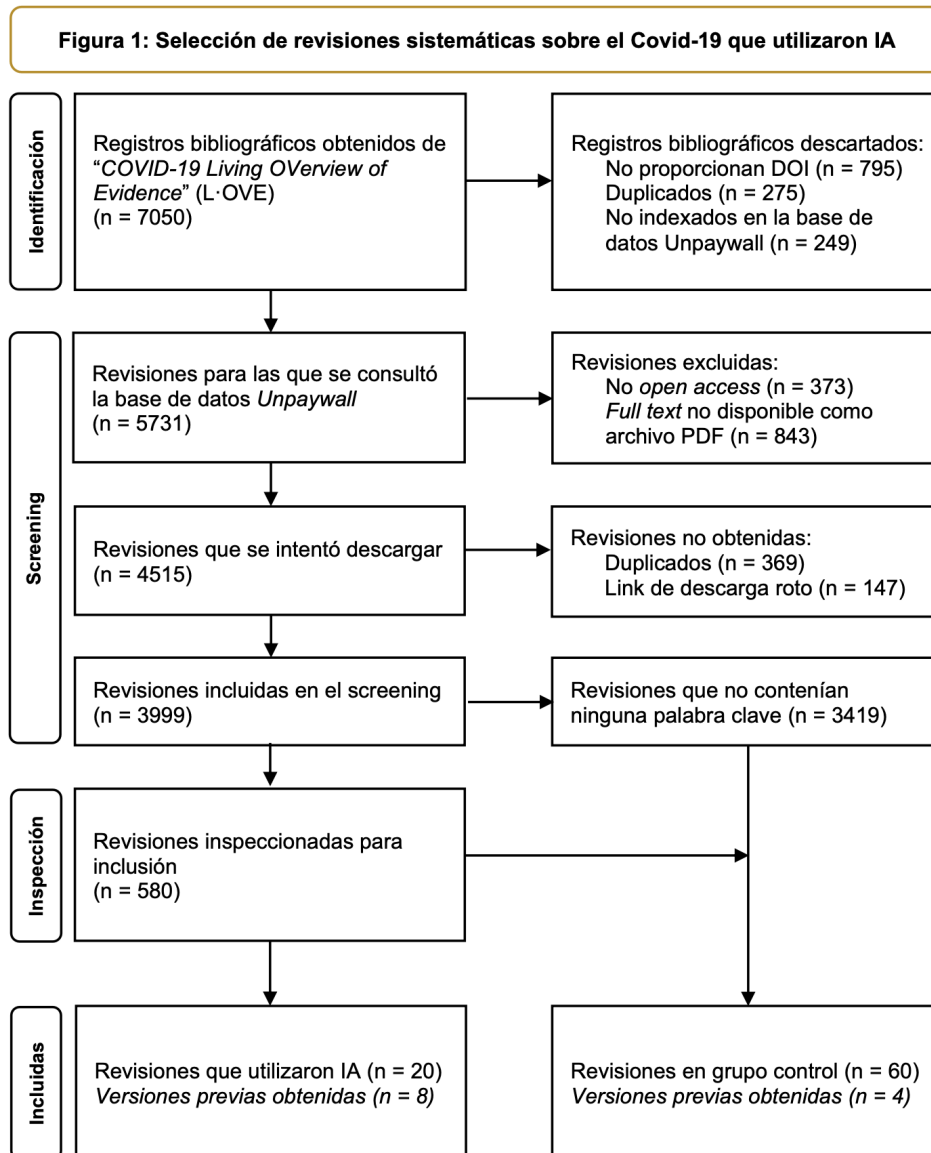


Figura 1. Diagrama de flujo de las revisiones sistemáticas identificadas, cribadas, inspeccionadas para elegibilidad e incluidas en nuestro estudio.

revisiones *rapid* (5/20 vs. 6/60, IC del 95 % -5,4 a 35,4 %, $P=0,092$), financiación (12/20 vs. 27/60, IC del 95 % -9,9 a 39,9 %, $P=0,245$) o estado de publicación (12/20 vs. 48/60, IC del 95 % -43,7 a 3,7 %, $P=0,074$). Los factores de impacto JCR en las revisiones publicadas en el grupo IA fueron significativamente mayores que los controles (mediana [IQR]: 8,9 [3,9-39,9] vs. 3,5 [2,6-5,5], $P<0,001$); en número de citas recibidas no mostraron diferencias (0,5 [0,0-13,5] vs. 0,6 [0,0-2,8], $P=0,600$).

En cuanto a las mediciones de la carga de trabajo, el grupo IA revisó más abstracts por autor (302,2 [126,7-804,3] frente a 140,3 [43,8-378,2], $P=0,009$) y por estudio incluido (189,0 [94,1-365,8] frente a 26,9 [13,7-64,1], $P<0,001$), mientras que inspeccionaron menos *full texts* por autor (5,3 [3,7-16,1] vs. 14,0 [6,5-37,2], $P=0,005$) y por estudio incluido (3,8 [2,4-5,3] vs. 3,4 [2,0-6,2], $P=0,481$).

No se observaron diferencias en los tiempos de elaboración de las revisiones

Tabla 1 Variables extraídas para las revisiones que usaron inteligencia artificial (IA) y los controles

Características	Grupo IA (n = 20)		Controles (n = 60)		Δ	χ^2	Valor P
	n	(%)	n	(%)			
Revisiones tipo <i>rapid</i>	5	(25 %)	6	(10 %)	15 %	2.846	0.092
Revisiones tipo <i>living</i>	5	(25 %)	3	(5 %)	20 %	6.667	0.010
Recibieron financiación	12	(60 %)	27	(45 %)	15 %	1.351	0.245
Publicadas	12	(60 %)	48	(80 %)	-20 %	3.2	0.074
	Mediana	IQR	Mediana	IQR		Wilcoxon W	Valor P
JCR Impact Factor de la revista	9	(4-40)	3	(3-6)		409	<0.001
Citas recibidas por mes	1	(0-13)	1	(0-3)		647	0.600
Abstracts inspeccionados							
por autor	302	(127-804)	140	(44-378)		1126	0.009
por estudio incluido	189	(94-366)	27	(14-64)		1443	<0.001
Full texts inspeccionados							
por autor	5	(4-16)	14	(7-37)		504.5	0.005
por estudio incluido	4	(2-5)	3	(2-6)		883.5	0.481
Tiempo de elaboración (días)	74	(48-118)	123	(53-221)		183.5	0.205
Puntuación AMSTAR-2 (sobre 13)	8	(5-9)	6	(4-8)		740.5	0.119

Se utilizó la prueba de chi-cuadrado de Pearson para comparar las proporciones de revisiones *rapid*, *living*, que recibieron financiación y que fueron publicadas, y el test de Wilcoxon-Mann-Whitney para el resto de comparaciones. Las medianas y rangos intercuartílicos (IQR, expresados como *cuartil 1 – cuartil 3*) se muestran redondeados al entero más cercano.

Δ : Diferencias absolutas en puntos porcentuales entre las revisiones que usaron IA y las de control. χ^2 : estadístico de la prueba de chi-cuadrado de Pearson. *Wilcoxon W*: estadístico del test de Wilcoxon-Mann-Whitney (rank-sum test).

prerregistradas (74,0 [47,5-117,5] frente a 123,0 [53,0-221,0], $P=0,205$). Las puntuaciones medias obtenidas en la evaluación metodológica AMSTAR-2 (Apéndice D) no fueron significativamente mayores en el grupo IA (7,5 [5,3-9,1] vs. 6,3 [3,9-8,0] puntos sobre 13, $P=0,119$), mostrando ambos grupos una alta heterogeneidad de resultados, como se puede observar en la *Figura 3*. En comparación con los controles, las revisiones con IA obtuvieron peores resultados en la pregunta 4 (estrategia de búsqueda de literatura, -12 %) y mejores en la pregunta 6 (extracción de datos por duplicado, 35 %), mientras que mostraron diferencias mínimas en la pregunta 5 (*screening* por duplicado, 7 %). Ambos grupos obtuvieron las puntuaciones más bajas en las preguntas 7 (proporcionar una lista de estudios excluidos) y 10 (informar sobre las fuentes de financiación de los estudios incluidos).

3.4. Descripción narrativa de los usos de la IA

Según la etapa del proceso de revisión en la que se utilizó IA, podemos clasificar las 20 revisiones del grupo IA en tres categorías, como se muestra en la *Tabla 2*.

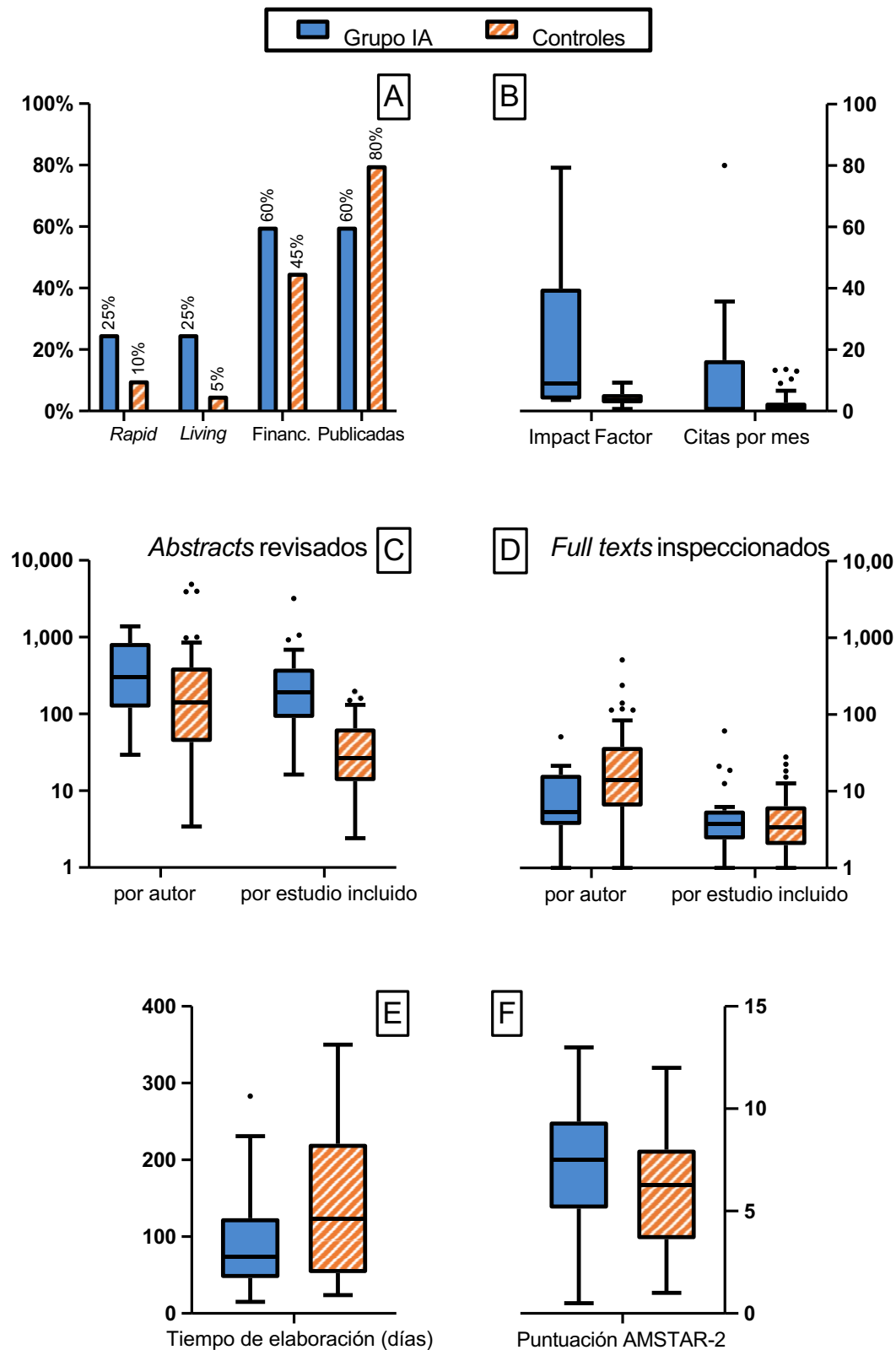


Figura 2. Características de las revisiones incluidas. Diagrama *box-and-whisker* (de “cajas y bigotes”: las cajas encierran los cuartiles Q1–Q3, sus líneas centrales representan la mediana, y los bigotes se extienden hasta los puntos de datos más lejanos dentro del intervalo de 1,5 IQR). El panel A compara la proporción de revisiones *rapid*, *living*, financiadas y publicadas entre los grupos; el panel B presenta los Factores de Impacto JCR 2020 de las revistas y los recuentos de citas por mes de cada grupo; los paneles C y D muestran las mediciones de la carga de trabajo de los autores: abstracts revisados y *full texts* inspeccionados, por autor y por estudio incluido; el panel E muestra los tiempos medios de elaboración (en días) de las revisiones de cada grupo; y el panel F representa sus evaluaciones metodológicas AMSTAR–2.

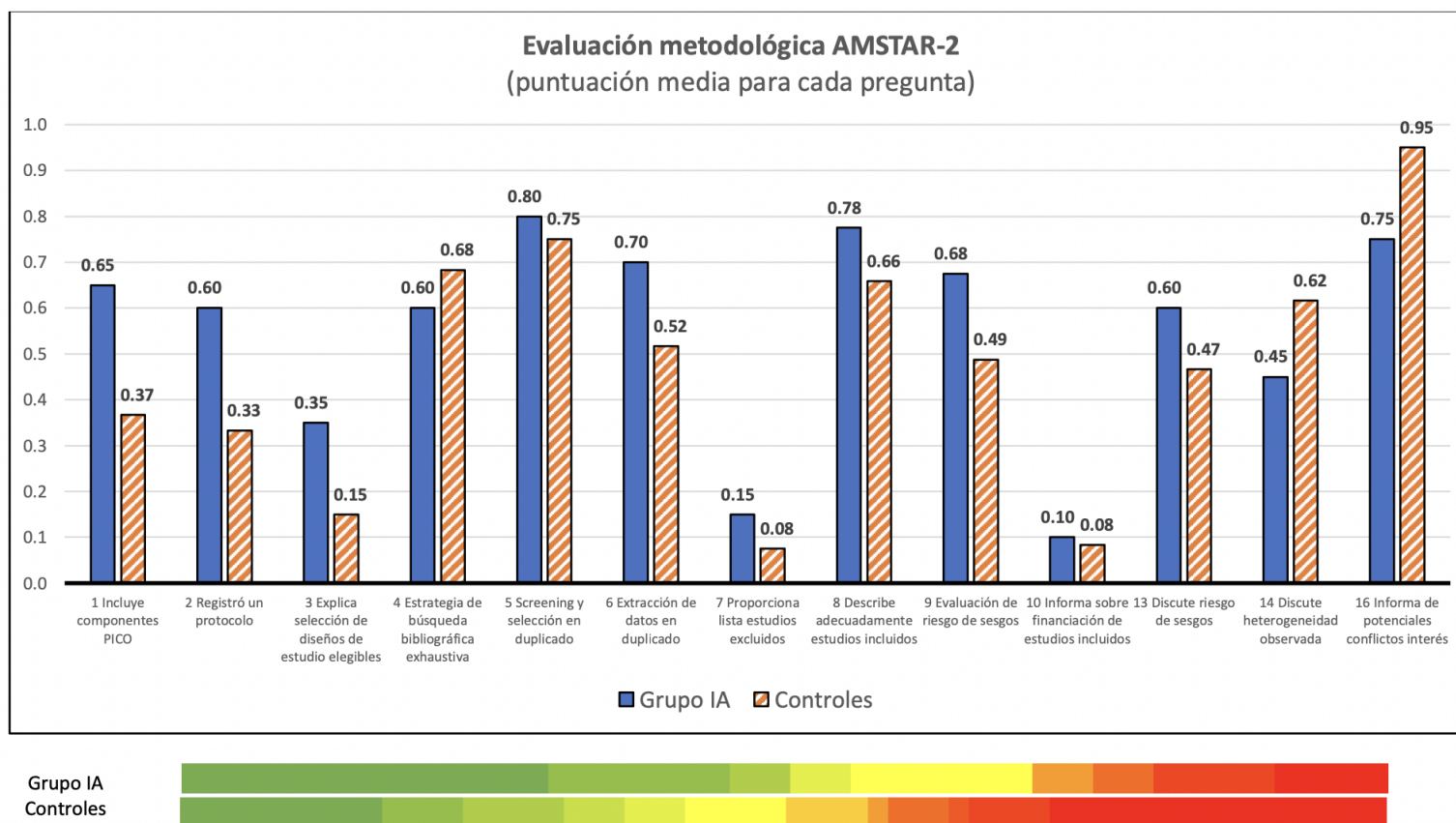


Figura 3. Resumen de la evaluación metodológica AMSTAR-2. El gráfico de la parte superior muestra las puntuaciones medias obtenidas en cada una de las preguntas evaluadas en el grupo de revisiones sistemáticas con uso de inteligencia artificial (IA) (barras azules) y en el grupo control (barras naranjas). Las líneas de colores de la parte inferior representan visualmente la heterogeneidad de resultados obtenidos en las revisiones sistemáticas de ambos grupos (el gradiente representa las puntuaciones obtenidas: rojo=4; amarillo=6,5; verde=9).

Tabla 2a Uso de IA: Asistencia en la búsqueda bibliográfica

Ref.	Título	Autor	Revista	Software utilizado	¿Código abierto?
[31]	Prevalence of Gastrointestinal Symptoms and Fecal Viral Shedding in Patients with Coronavirus Disease 2019	<i>Parasa et al.</i>	JAMA Network Open	CORD-19	Parcial
[32]	The influence of comorbidity on the severity of COVID-19 disease: systematic review and analysis	<i>Zaki et al.</i>	<i>preprint</i>	CORD-19 + Okapi BM25	Sí
[33]	The Estimations of the COVID-19 Incubation Period: A Scoping Reviews of the Literature	<i>Zaki et al.</i>	Journal of Infection and Public Health	CORD-19 + BioBERT	Sí
[34]	Ocular toxicity and Hydroxychloroquine: A Rapid Meta-Analysis	<i>Michelson et al.</i>	<i>preprint</i>	GenesisAI	No
[35]	A Systematic Review of the Incubation Period of SARS-CoV-2: The Effects of Age, Biological Sex, and Location on Incubation Period	<i>Daley et al.</i>	<i>preprint</i>	No reportado	No

Tabla 2b Uso de IA: Filtrado de ensayos controlados aleatorizados

Ref.	Título	Autor	Revista	Software utilizado	¿Código abierto?
[36]	Impact of remdesivir on 28 day mortality in hospitalized patients with COVID-19: February 2021 Meta-analysis	<i>Robinson et al.</i>	<i>preprint</i>	RobotSearch	Sí
[37]	Impact of systemic corticosteroids on hospitalized patients with COVID-19: January 2021 Meta-analysis of randomized controlled trials	<i>Robinson et al.</i>	<i>preprint</i>	RobotSearch	Sí
[25]	Prophylaxis against COVID-19: living systematic review and network meta-analysis	<i>Bartoszko et al.</i>	BMJ	RobotSearch	Sí
[26]	Drug treatments for COVID-19: living systematic review and network meta-analysis	<i>Siemieniuket al.</i>	BMJ	RobotSearch	Sí
[38]	Adverse effects of remdesivir, hydroxychloroquine, and lopinavir/ritonavir when used for COVID-19: systematic review and meta-analysis of randomized trials	<i>Izcovich et al.</i>	<i>preprint</i>	RobotSearch	Sí
[39]	Tocilizumab and sarilumab alone or in combination with corticosteroids for COVID-19: A systematic review and network meta-analysis	<i>Zeraatkar et al.</i>	<i>preprint</i>	RobotSearch	Sí
[40]	Clinical trials in COVID-19 management & prevention: A meta-epidemiological study examining methodological quality	<i>Honarmand et al.</i>	Journal of Clinical Epidemiology	RobotSearch	Sí

Tabla 2c Uso de IA: Automatización del screening

Ref.	Título	Autor	Revista	Software utilizado	¿Código abierto?
[41]	Impacts of school closures on physical and mental health of children and young people: a systematic review	Viner <i>et al.</i>	<i>preprint</i>	EPPI Reviewer	No
[27]	Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal	Wynants <i>et al.</i>	BMJ	EPPI Reviewer	No
[28]	Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection (Review)	Dinnes <i>et al.</i>	Cochrane Database of Systematic Reviews	EPPI Reviewer	No
[29]	Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19	Struyf <i>et al.</i>	Cochrane Database of Systematic Reviews	EPPI Reviewer	No
[42]	Are medical procedures that induce coughing or involve respiratory suctioning associated with increased generation of aerosols and risk of SARS-CoV-2 infection? A rapid systematic review	Wilson <i>et al.</i>	Journal of Hospital Infection	EPPI Reviewer	No
[43]	Risk and Protective Factors in the COVID-19 Pandemic: A Rapid Evidence Map	Elmore <i>et al.</i>	Frontiers in Public Health	SWIFT-Active Screener	No
[44]	Tocilizumab and Systemic Corticosteroids in the Management of COVID-19 Patients: A Systematic Review and Meta-Analysis	Alkofide <i>et al.</i>	International Journal of Infectious Diseases	Abstrackr	Sí
[30]	Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis	Chu <i>et al.</i>	The Lancet	Evidence Prime	No

Tabla 2. Tabla que muestra las diferentes herramientas de inteligencia artificial (IA) que se utilizaron en la elaboración de revisiones sistemáticas sobre COVID-19 (con enlaces a aquellas de código abierto), según su ámbito de aplicación: asistencia en la búsqueda bibliográfica, filtrado de ensayos controlados aleatorizados (ECA) y automatización del *screening*.

Asistencia en la búsqueda bibliográfica

Tres revisiones [31, 32, 33] complementaron sus procedimientos de búsqueda con consultas de tipo “pregunta abierta” en CORD-19 [45], una base de datos sobre el COVID-19 estructurada para facilitar el uso de sistemas de minería de textos y *deep-learning*: Zaki *et al.* [32] utilizaron un repositorio de GitHub basado en el algoritmo de búsqueda *Okapi BM25*; Zaki *et al.* [33] emplearon *BioBERT*, un sistema revisado por pares [46] y de código abierto preentrenado para el análisis de literatura biomédica; y Parasa *et al.* [31] no proporcionaron detalles sobre el motor de búsqueda empleado. Además, Michelson *et al.* [34] utilizaron software privado de la empresa *GenesisAI* para producir un “*rapid meta-analysis*” como prueba de concepto de su producto. Daley *et al.* [35] no proporcionaron ninguna información sobre el software empleado. Sólo 2 revisiones en este subgrupo se encontraban publicadas, y ninguna pre-registró un protocolo. La puntuación media en AMSTAR-2 fue de 3,7/13.

Filtrado de ensayos controlados aleatorizados

Siete artículos [25, 26, 36, 37, 38, 39, 40] emplearon *RobotSearch*, un software revisado por pares [47] y de código abierto para identificar, de entre las referencias bibliográficas proporcionadas por el usuario, aquellas correspondientes a ensayos controlados aleatorizados (ECA). Se basa en una red neuronal entrenada con datos de las revisiones de Cochrane y destaca por su facilidad de uso (no requiere instalación) y flexibilidad (ya que permite diferentes niveles de sensibilidad, incluido uno destinado específicamente a revisiones sistemáticas, así como la integración con otros *scripts*).

En nuestra muestra, *RobotSearch* se utilizó particularmente con frecuencia en revisiones *living* o parcialmente automatizadas. Dos de las revisiones que utilizaron *RobotSearch* fueron la de Bartoszko *et al.* [25] un meta-análisis en red sobre la profilaxis del COVID-19, y Siemieniuk *et al.* [26], un *living meta-analysis* de ensayos aleatorizados para guiar las recomendaciones de la Organización Mundial de la Salud (OMS) sobre la terapéutica del COVID-19, de los cuales Izcovich *et al.* [38] y Zeraatkar *et al.* [39] son subestudios separados. Ambos forman parte del proyecto “*BMJ Rapid Recommendations*” y publican sus conclusiones y análisis preliminares en un sitio web en constante actualización. La puntuación media en AMSTAR-2 fue de 7,5/13.

Automatización del screening

Encontramos ocho artículos [27, 28, 29, 30, 41, 42, 43, 44] que hacían uso de procedimientos de *screening* potenciados por IA. Cinco de ellos [27, 28, 29, 41, 42] utilizaron *EPPI Reviewer*, una plataforma web (de acceso tipo *shareware*) para asistir en la elaboración de todo tipo de revisiones sistemáticas. Ofrece una gran variedad de funciones, desde la gestión bibliográfica hasta facilitar el trabajo colaborativo, así como la capacidad de sugerir otros estudios relevantes, la agrupación automática de

artículos y la minería de textos. En particular, las revisiones incluidas utilizaron su módulo “*SGDClassifier*” para priorizar el *screening* de artículos con más probabilidad de ser incluidos de acuerdo con las decisiones previas del usuario. Como resultado, tanto Wynants *et al.* [27] como dos revisiones Cochrane [28, 29] citan una reducción del 80 % en la carga de *screening* gracias a esta herramienta.

Otras dos revisiones utilizaron técnicas de automatización del *screening* englobadas en plataformas de asistencia a la elaboración de revisiones sistemáticas: *SWIFT-Active Screener* [48] en Elmore *et al.* [43], configurado para asegurar un recall mínimo (porcentaje de captura de estudios relevantes) como criterio de detención del *screening*; y *Evidence Prime* de Chu *et al.* [30] para verificar y complementar el *screening* manual. Por último, Alkofide *et al.* [44] utilizaron *Abstrackr*, el único software de código abierto en esta categoría, que utiliza el feedback de los artículos previamente seleccionados y rechazados por el usuario para guiar el proceso de *screening*. Las evaluaciones de esta herramienta publicadas en la literatura [49] sugieren un gran ahorro de trabajo en la producción de revisiones sistemáticas a costa de una tasa de falsos negativos del 0,1 %.

Entre las revisiones analizadas en este estudio, este subgrupo presentó las puntuaciones más altas en la herramienta de valoración AMSTAR-2 (9,1/13), destacando las menciones de dos revisiones Cochrane [28, 29] (12 puntos) y un meta-análisis rápido [30] publicado en *The Lancet* (10,5 puntos). A diferencia de las revisiones de las otras categorías, que priorizaron la profundidad de la búsqueda, el uso de herramientas basadas en IA en este subgrupo estuvo motivado por la gran carga en *screening* a la que se enfrentaron los revisores: citando a Dinnes *et al.* [28], «*se necesitaba un enfoque más eficiente para procesar la creciente producción científica sobre el COVID-19*».

4. DISCUSIÓN

En este trabajo se evaluó si los potenciales beneficios de implementar IA en la elaboración de revisiones sistemáticas se han visto reflejados en las revisiones sobre el COVID-19. Encontramos que la IA se utilizó raramente, apareciendo sólo en el 0,7 % de las revisiones estudiadas, pero que se asoció significativamente con una reducción de la carga de trabajo en *screening* de los autores y la publicación en revistas con mayor factor de impacto. Ser una revisión *living* se asoció con el uso de IA, siendo los usos más comunes la optimización del *screening* (priorizando los estudios con alta probabilidad de ser relevantes) y la selección de ensayos controlados aleatorizados.

Como limitación de nuestro estudio, destacamos su baja potencia estadística debido al escaso número de revisiones que utilizaron IA. Anticipando la limitada disponibilidad de revisiones con IA, adoptamos un procedimiento de *screening* muy sensible, procesando más de 7 000 referencias bibliográficas de revisiones sistemáticas sobre el COVID-19 (combinando la consulta de expertos para la selección de

palabras clave con un potente motor de búsqueda), y elegimos un ratio de 3:1 para el tamaño del grupo de control para minimizar el riesgo de errores estadísticos de tipo II. El uso de *L-OVE* como base de datos principal permitió acceder a múltiples fuentes bibliográficas relevantes y actualizadas de forma sistemática y automatizable; sin embargo, nuestra estrategia de búsqueda podría mostrar una menor sensibilidad para los informes institucionales y *whitepapers*, que a menudo no son indexados por las bases de datos tradicionales. El potencial impacto de errores puntuales en la descarga de revisiones y la exclusión de las revisiones no *open-access* de nuestro estudio es incierto; su impacto en la generalización de nuestros resultados debe interpretarse teniendo en cuenta la gran diversidad de fuentes secundarias accesibles a través de *L-OVE* y la alta accesibilidad a la investigación sobre el COVID-19 durante la pandemia. Además, el uso de las fechas de publicación como variable de emparejamiento permitió una selección de controles guiada por nuestro *script* (para minimizar el riesgo de sesgos), pero impidió el uso de otras variables de control deseables, como el tamaño o el enfoque temático de las revisiones.

Debemos resaltar que el cálculo de la carga de *screening* “por autor” en lugar de “por revisor que participa en el cribado” puede infraestimar estas variables en revisiones con equipos grandes (cuando no todos sus autores participan en el cribado). Tener mayor número de autores también podría relacionarse con la disponibilidad de recursos de los equipos, y por tanto con el acceso a asesoramiento y expertos en IA. Asimismo, los grupos con más recursos y con apoyo de expertos en IA podrían contar con mayores facilidades de acceso a revistas bien indexadas, lo que podría sesgar los análisis del factor de impacto a favor de la IA. Por otra parte, la evaluación con AMSTAR-2 se realizó inevitablemente sin poder cegar a los evaluadores en cuanto al uso o no de IA, lo cual, dada la subjetividad de ciertos aspectos de la evaluación metodológica, podría haber influido en sus resultados. Por último, el uso de los recuentos de citas para medir el impacto de las revisiones es un abordaje imperfecto con deficiencias conocidas, como el “*citation bias*” (los resultados estadísticamente significativos suelen citarse más que otros) o la propia autoridad de los autores [50] y puede particularmente subestimar el impacto de las revisiones publicadas más recientemente.

En promedio, un equipo de 5 revisores tarda 15 meses en completar una revisión sistemática tradicional [51] con tasas de errores estimadas en torno al 10% [52]. Enfrentarse a la pandemia del COVID-19 exigía contar con revisiones sistemáticas sólidas con urgencia, ya que cualquier retraso suponía un coste tanto en términos de vidas perdidas como en daños económicos. Sin embargo, a pesar del crecimiento exponencial que han experimentado los campos de la IA y el *machine-learning* durante los últimos años, estos desempeñaron un papel sorprendentemente limitado en las revisiones sistemáticas sobre el COVID-19. Nuestras conclusiones concuerdan con las de trabajos anteriores [14] que sugieren que los beneficios que la IA puede proporcionar en la realización de revisiones sistemáticas son desconocidos para la mayoría de los revisores, y la relativa heterodoxia de sus métodos podría dificultar

inicialmente su aceptación por parte de la comunidad científica. El software de código abierto, generalmente más propenso a ser adoptado en dichos círculos, jugará un papel esencial en este aspecto.

Nuestra descripción narrativa de las revisiones incluidas en este estudio mostró que ninguna hizo uso simultáneo de más de una herramienta de IA. Un enfoque más cohesivo, que integre la IA en cada paso del proceso de revisión, ahorraría tiempo a los revisores que desearan interconectar diferentes herramientas con formatos a veces incompatibles. Las herramientas de asistencia semiautomatizadas fueron una de las áreas donde la IA mostró una mayor adopción, y la variedad de opciones de software (como *EPPI Reviewer*, ya adoptado como herramienta oficial de producción de revisiones Cochrane) fue mayor. Por el contrario, la automatización completa sólo fue empleada por *RobotSearch* (una herramienta para filtrar ensayos aleatorizados ampliamente evaluada en la literatura), lo que sugiere que la adopción de soluciones cada vez más automatizadas puede requerir la producción previa de evaluaciones más detalladas de sus costes potenciales (tanto en pérdida de artículos como en riesgo de sesgos) en balance con sus aportaciones en productividad.

Conclusión

La necesidad de automatización en las síntesis de evidencia es obvia, ya que la carga de trabajo de los revisores crece tan rápido como las ciencias biomédicas. La adopción de nuevas tecnologías puede llevar tiempo, pero aprovechar el potencial de la IA en la elaboración de revisiones sistemáticas debe ser una prioridad. En el futuro, la IA debe incorporarse a las revisiones sistemáticas como el siguiente paso para facilitar la toma de decisiones de manera más basada en la evidencia, rápida y precisa.

Referencias

1. Lasserson T, Thomas J, and Higgins J. Starting a review. *Cochrane handbook for systematic reviews of interventions*. 6.3. Cochrane, 2022. Chap. 1. Available from: training.cochrane.org/handbook
2. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, and Coiera E. Systematic review automation technologies. *Systematic Reviews* 2014; 3:74. DOI: [10.1186/2046-4053-3-74](https://doi.org/10.1186/2046-4053-3-74)
3. Nussbaumer-Streit B, Ellen M, Klerings I, et al. Resource use during systematic review production varies widely: a scoping review. *Journal of Clinical Epidemiology* 2021; 139:287–96. DOI: [10.1016/j.jclinepi.2021.05.019](https://doi.org/10.1016/j.jclinepi.2021.05.019)
4. Gill D, Baker EH, and Hitchings AW. We need clinical guidelines fit for a pandemic. *BMJ* 2021; 373:n1093. DOI: [10.1136/bmj.n1093](https://doi.org/10.1136/bmj.n1093)
5. Knottnerus JA and Tugwell P. Methodological challenges in studying the COVID-19 pandemic crisis. *Journal of Clinical Epidemiology* 2020; 121:A5–A7. DOI: [10.1016/j.jclinepi.2020.04.001](https://doi.org/10.1016/j.jclinepi.2020.04.001)
6. Li Y, Cao L, Zhang Z, et al. Reporting and methodological quality of COVID-19 systematic reviews needs to be improved: an evidence mapping. *Journal of Clinical Epidemiology* 2021; 135:17–28. DOI: [10.1016/j.jclinepi.2021.02.021](https://doi.org/10.1016/j.jclinepi.2021.02.021)
7. Jung RG, Di Santo P, Clifford C, et al. Methodological quality of COVID-19 clinical research. *Nature Communications* 2021; 12:943. DOI: [10.1038/s41467-021-21220-5](https://doi.org/10.1038/s41467-021-21220-5)
8. Tricco AC, Garritty CM, Boulos L, et al. Rapid review methods more challenging during COVID-19: commentary with a focus on 8 knowledge synthesis steps. *Journal of Clinical Epidemiology* 2020; 126:177–83. DOI: [10.1016/j.jclinepi.2020.06.029](https://doi.org/10.1016/j.jclinepi.2020.06.029)
9. Biesty L, Meskell P, Glenton C, et al. A QuESt for speed: rapid qualitative evidence syntheses as a response to the COVID-19 pandemic. *Systematic Reviews* 2020; 9:256. DOI: [10.1186/s13643-020-01512-5](https://doi.org/10.1186/s13643-020-01512-5)
10. Amezcua-Prieto C, Fernández-Luna JM, Huete-Guadix JF, Bueno-Cavanillas A, and Khan KS. Artificial intelligence and automation of systematic reviews in women’s health. *Current opinion in Obstetrics & Gynecology* 2020; 32:335–41. DOI: [10.1097/GCO.0000000000000643](https://doi.org/10.1097/GCO.0000000000000643)
11. O’Mara-Eves A, Thomas J, McNaught J, Miwa M, and Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic reviews* 2015; 4:5. DOI: [10.1186/2046-4053-4-5](https://doi.org/10.1186/2046-4053-4-5)
12. Marshall IJ and Wallace BC. Toward systematic review automation: A practical guide to using machine learning tools in research synthesis. *Systematic Reviews* 2019; 8:163. DOI: [10.1186/s13643-019-1074-9](https://doi.org/10.1186/s13643-019-1074-9)
13. Thomas J, Noel-Storr A, Marshall I, et al. Living systematic reviews: 2. Combining human and machine effort. *Journal of Clinical Epidemiology* 2017; 91:31–7. DOI: [10.1016/j.jclinepi.2017.08.011](https://doi.org/10.1016/j.jclinepi.2017.08.011)
14. Scott AM, Forbes C, Clark J, Carter M, Glasziou P, and Munn Z. Systematic review automation tools improve efficiency but lack of knowledge impedes their adoption: a survey. *Journal of Clinical Epidemiology* 2021; 138:80–94. DOI: [10.1016/j.jclinepi.2021.06.030](https://doi.org/10.1016/j.jclinepi.2021.06.030)
15. Thomas J, McDonald S, Noel-Storr A, et al. Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews. *Journal of Clinical Epidemiology* 2021; 133:140–51. DOI: [10.1016/j.jclinepi.2020.11.003](https://doi.org/10.1016/j.jclinepi.2020.11.003)
16. Schmidt L, Olorisade BK, McGuinness LA, Thomas J, and Higgins JP. Data extraction methods for systematic review (semi)automation: A living systematic review. *F1000Research* 2021; 10:401. DOI: [10.12688/f1000research.51117.1](https://doi.org/10.12688/f1000research.51117.1)
17. Jonnalagadda SR, Goyal P, and Huffman MD. Automating data extraction in systematic reviews: A systematic review. *Systematic Reviews* 2015; 4:78. DOI: [10.1186/s13643-015-0066-7](https://doi.org/10.1186/s13643-015-0066-7)

18. Mbuagbaw L, Lawson DO, Puljak L, Allison DB, and Thabane L. A tutorial on methodological studies: The what, when, how and why. *BMC Medical Research Methodology* 2020; 20:226. DOI: [10.1186/s12874-020-01107-7](https://doi.org/10.1186/s12874-020-01107-7)
19. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 2021; 372:n71. DOI: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)
20. Tercero-Hidalgo JR, Khan KS, Bueno-Cavanillas A, et al. Covid-19 systematic evidence synthesis with Artificial Intelligence: a Review of Reviews. *Open Science Forum Registries* 2021. DOI: [10.17605/OSF.IO/H5DAW](https://doi.org/10.17605/OSF.IO/H5DAW)
21. Tercero-Hidalgo JR, Khan KS, Bueno-Cavanillas A, et al. COVID-19 evidence syntheses with artificial intelligence: an empirical study of systematic reviews. *Dryad Dataset* 2021. DOI: [10.5061/dryad.9kd51c5j6](https://doi.org/10.5061/dryad.9kd51c5j6)
22. Rada G, Verdugo-Paiva F, Ávila C, et al. Evidence synthesis relevant to COVID-19: a protocol for multiple systematic reviews and overviews of systematic reviews. *eng. Medwave* 2020; 20:e7868. DOI: [10.5867/medwave.2020.03.7867](https://doi.org/10.5867/medwave.2020.03.7867)
23. Dhakal K. Unpaywall. *Journal of the Medical Library Association: JMLA* 2019; 107:286–8. DOI: [10.5195/jmla.2019.650](https://doi.org/10.5195/jmla.2019.650)
24. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017; 358:j4008. DOI: [10.1136/bmj.j4008](https://doi.org/10.1136/bmj.j4008)
25. Bartoszko JJ, Siemieniuk RA, Kum E, et al. Prophylaxis against covid-19: Living systematic review and network meta-analysis. *BMJ* 2021; 373:n949. DOI: [10.1136/bmj.n949](https://doi.org/10.1136/bmj.n949)
26. Siemieniuk RAC, Bartoszko JJ, Ge L, et al. Drug treatments for covid-19: Living systematic review and network meta-Analysis. *BMJ* 2020; 370:m2980. DOI: [10.1136/bmj.m2980](https://doi.org/10.1136/bmj.m2980)
27. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ* 2020; 369:m1328. DOI: [10.1136/bmj.m1328](https://doi.org/10.1136/bmj.m1328)
28. Dinnes J, Deeks JJ, Berhane S, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database of Systematic Reviews* 2021 :CD013705. DOI: [10.1002/14651858.CD013705.pub2](https://doi.org/10.1002/14651858.CD013705.pub2)
29. Struyf T, Deeks JJ, Dinnes J, et al. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database of Systematic Reviews* 2021 :CD013665. DOI: [10.1002/14651858.CD013665.pub2](https://doi.org/10.1002/14651858.CD013665.pub2)
30. Chu DK, Akl EA, Duda S, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet* 2020; 395:1973–87. DOI: [10.1016/S0140-6736\(20\)31142-9](https://doi.org/10.1016/S0140-6736(20)31142-9)
31. Parasa S, Desai M, Thoguluva Chandrasekar V, et al. Prevalence of Gastrointestinal Symptoms and Fecal Viral Shedding in Patients with Coronavirus Disease 2019: A Systematic Review and Meta-analysis. *JAMA Network Open* 2020; 3:e2011335. DOI: [10.1001/jamanetworkopen.2020.11335](https://doi.org/10.1001/jamanetworkopen.2020.11335)
32. Zaki N, Mohamed EA, Ibrahim S, and Khan G. The influence of comorbidity on the severity of COVID-19 disease: systematic review and analysis. *medRxiv* 2020 :2020.06.18.20134478. DOI: [10.1101/2020.06.18.20134478](https://doi.org/10.1101/2020.06.18.20134478)
33. Zaki N and Mohamed EA. The estimations of the COVID-19 incubation period: A scoping reviews of the literature. *Journal of Infection and Public Health* 2021; 14:638–46. DOI: [10.1016/j.jiph.2021.01.019](https://doi.org/10.1016/j.jiph.2021.01.019)
34. Michelson M, Chow T, Martin N, Ross M, Tee A, and Minton S. Ocular toxicity and Hydroxychloroquine: A Rapid Meta-Analysis. *medRxiv* 2020 :2020.04.28.20083378. DOI: [10.1101/2020.04.28.20083378](https://doi.org/10.1101/2020.04.28.20083378)
35. Daley C, Fydenkevez M, and Ackerman-Morris S. A Systematic Review of the Incubation Period of SARS-CoV-2: The Effects of Age, Biological Sex, and Location on Incubation Period. *medRxiv* 2020 :2020.12.23.20248790. DOI: [10.1101/2020.12.23.20248790](https://doi.org/10.1101/2020.12.23.20248790)
36. Robinson R, Prakash V, Tamimi RA, et al. Impact of remdesivir on 28 day mortality in hospitalized patients with COVID-19: February 2021 Meta-analysis. *medRxiv* 2021 :2021.03.04.21252903. DOI: [10.1101/2021.03.04.21252903](https://doi.org/10.1101/2021.03.04.21252903)

37. Robinson R, Prakash V, Tamimi RA, Albast N, and Al-Bast B. Impact of systemic corticosteroids on hospitalized patients with COVID-19: January 2021 Meta-analysis of randomized controlled trials. *medRxiv* 2021 :2021.02.03.21251065. DOI: [10.1101/2021.02.03.21251065](https://doi.org/10.1101/2021.02.03.21251065)
38. Izcovich A, Siemieniuk RA, Bartoszko JJ, et al. Adverse effects of remdesivir, hydroxychloroquine and lopinavir/ritonavir when used for COVID-19: systematic review and meta-analysis of randomised trials. *BMJ Open* 2022; 12:e048502. DOI: [10.1136/bmjopen-2020-048502](https://doi.org/10.1136/bmjopen-2020-048502)
39. Zeraatkar D, Cusano E, Martínez JPD, et al. Use of tocilizumab and sarilumab alone or in combination with corticosteroids for covid-19: systematic review and network meta-analysis. *BMJ Medicine* 2022; 1:e000036. DOI: [10.1101/2021.07.05.21259867](https://doi.org/10.1101/2021.07.05.21259867)
40. Honarmand K, Penn J, Agarwal A, et al. Clinical trials in COVID-19 management & prevention: A meta-epidemiological study examining methodological quality. *Journal of Clinical Epidemiology* 2021; 139:68–79. DOI: [10.1016/j.jclinepi.2021.07.002](https://doi.org/10.1016/j.jclinepi.2021.07.002)
41. Viner R, Russell S, Saullé R, et al. Impacts of school closures on physical and mental health of children and young people: a systematic review. *medRxiv* 2021 :2021.02.10.21251526. DOI: [10.1101/2021.02.10.21251526](https://doi.org/10.1101/2021.02.10.21251526)
42. Wilson J, Carson G, Fitzgerald S, et al. Are medical procedures that induce coughing or involve respiratory suctioning associated with increased generation of aerosols and risk of SARS-CoV-2 infection? A rapid systematic review. *Journal of Hospital Infection* 2021; 116:37–46. DOI: [10.1016/j.jhin.2021.06.011](https://doi.org/10.1016/j.jhin.2021.06.011)
43. Elmore R, Schmidt L, Lam J, et al. Risk and Protective Factors in the COVID-19 Pandemic: A Rapid Evidence Map. *Frontiers in Public Health* 2020; 8:582205. DOI: [10.3389/fpubh.2020.582205](https://doi.org/10.3389/fpubh.2020.582205)
44. Alkofide H, Almohaizeie A, Almuahini S, Alotaibi B, and Alkharfy KM. Tocilizumab and Systemic Corticosteroids in the Management of Patients with COVID-19: A Systematic Review and Meta-Analysis. *International Journal of Infectious Diseases* 2021; 110:320–9. DOI: [10.1016/j.ijid.2021.07.021](https://doi.org/10.1016/j.ijid.2021.07.021)
45. Lu Wang L, Lo K, Chandrasekhar Y, et al. COVID-19: The Covid-19 Open Research Dataset. *ArXiv* 2020 :2004.10706v4. DOI: [10.48550/arXiv.2004.10706](https://doi.org/10.48550/arXiv.2004.10706)
46. Lee J, Yoon W, Kim S, et al. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020; 36:1234–40. DOI: [10.1093/bioinformatics/btz682](https://doi.org/10.1093/bioinformatics/btz682)
47. Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, and Wallace BC. Machine learning for identifying Randomized Controlled Trials: An evaluation and practitioner’s guide. *Research Synthesis Methods* 2018; 9:602–14. DOI: [10.1002/jrsm.1287](https://doi.org/10.1002/jrsm.1287)
48. Howard BE, Phillips J, Tandon A, et al. SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. *Environment International* 2020; 138:105623. DOI: [10.1016/j.envint.2020.105623](https://doi.org/10.1016/j.envint.2020.105623)
49. Gates A, Johnson C, and Hartling L. Technology-assisted title and abstract screening for systematic reviews: A retrospective evaluation of the Abstrackr machine learning tool. *Systematic Reviews* 2018; 7:45. DOI: [10.1186/s13643-018-0707-8](https://doi.org/10.1186/s13643-018-0707-8)
50. Urrutia MJ, Duyx B, Swaen GM, Bouter LM, and Zeegers MP. Citation bias and other determinants of citation in biomedical research: findings from six citation networks. *Journal of Clinical Epidemiology* 2021; 132:71–8. DOI: [10.1016/j.jclinepi.2020.11.019](https://doi.org/10.1016/j.jclinepi.2020.11.019)
51. Borah R, Brown AW, Capers PL, and Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open* 2017; 7:e012545. DOI: [10.1136/bmjopen-2016-012545](https://doi.org/10.1136/bmjopen-2016-012545)
52. Wang Z, Nayfeh T, Tetzlaff J, O’Blenis P, and Murad MH. Error rates of human reviewers during abstract screening in systematic reviews. *PLoS ONE* 2020; 15:e0227742. DOI: [10.1371/journal.pone.0227742](https://doi.org/10.1371/journal.pone.0227742)

Apéndice A

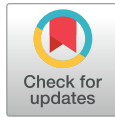
Artículo original

Journal of Clinical Epidemiology 148 (2022) 124-134

DOI [10.1016/j.jclinepi.2022.04.027](https://doi.org/10.1016/j.jclinepi.2022.04.027)



ELSEVIER



Journal of Clinical Epidemiology 148 (2022) 124–134

**Journal of
Clinical
Epidemiology**

ORIGINAL ARTICLE

Artificial intelligence in COVID-19 evidence syntheses was underutilized, but impactful: a methodological study

Juan R. Tercero-Hidalgo^{a,b,c,*}, Khalid S. Khan^{a,b}, Aurora Bueno-Cavanillas^{a,b,c},
Rodrigo Fernández-López^a, Juan F. Huete^d, Carmen Amezcua-Prieto^{a,b,c}, Javier Zamora^{b,e,f},
Juan M. Fernández-Luna^d

^aDepartment of Preventive Medicine and Public Health, University of Granada, Granada, Spain

^bCIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain

^cInstituto Biosanitario Granada (IBS-Granada), Granada, Spain

^dDepartment of Computer Science and Artificial Intelligence, School of Technology and Telecommunications Engineering, University of Granada, Granada, Spain

^eClinical Biostatistics Unit, Hospital Ramon y Cajal (IRYCIS), Madrid, Spain

^fInstitute for Metabolism and Systems Research, University of Birmingham, Birmingham, United Kingdom

Accepted 28 April 2022; Published online 2 May 2022

Abstract

Objectives: A rapidly developing scenario like a pandemic requires the prompt production of high-quality systematic reviews, which can be automated using artificial intelligence (AI) techniques. We evaluated the application of AI tools in COVID-19 evidence syntheses.

Study Design: After prospective registration of the review protocol, we automated the download of all open-access COVID-19 systematic reviews in the COVID-19 Living Overview of Evidence database, indexed them for AI-related keywords, and located those that used AI tools. We compared their journals' JCR Impact Factor, citations per month, screening workloads, completion times (from pre-registration to preprint or submission to a journal) and AMSTAR-2 methodology assessments (maximum score 13 points) with a set of publication date matched control reviews without AI.

Results: Of the 3,999 COVID-19 reviews, 28 (0.7%, 95% CI 0.47–1.03%) made use of AI. On average, compared to controls ($n = 64$), AI reviews were published in journals with higher Impact Factors (median 8.9 vs. 3.5, $P < 0.001$), and screened more abstracts per author (302.2 vs. 140.3, $P = 0.009$) and per included study (189.0 vs. 365.8, $P < 0.001$) while inspecting less full texts per author (5.3 vs. 14.0, $P = 0.005$). No differences were found in citation counts (0.5 vs. 0.6, $P = 0.600$), inspected full texts per included study (3.8 vs. 3.4, $P = 0.481$), completion times (74.0 vs. 123.0, $P = 0.205$) or AMSTAR-2 (7.5 vs. 6.3, $P = 0.119$).

Conclusion: AI was an underutilized tool in COVID-19 systematic reviews. Its usage, compared to reviews without AI, was associated with more efficient screening of literature and higher publication impact. There is scope for the application of AI in automating systematic reviews. © 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Artificial intelligence; Systematic review; COVID-19; Automation; Research design; Bibliometrics

Funding and conflicts of interest information: JRTH thanks the University of Granada for supporting his work through a Research Initiation Grant for Undergraduate Students. There are no conflicts of interest to declare.

Data statement: The data that support the findings of this study are available in Dryad (DOI [10.5061/dryad.9kd51c5j6](https://doi.org/10.5061/dryad.9kd51c5j6)). These data were derived from the following resources available in the public domain: [Living Overview of the Evidence](#), [Unpaywall](#), 2020 [Journal Citation Reports](#). The protocol of this study was pre-registered at Open Science Forum Registries (DOI [10.17605/OSF.IO/H5DAW](https://doi.org/10.17605/OSF.IO/H5DAW)).

CRedit author statement: **Juan R. Tercero-Hidalgo:** Methodology, Software, Investigation, Writing – Original Draft. **Khalid S. Khan:** Conceptualization, Methodology, Writing – Review & Editing, Supervision. **Aurora Bueno-Cavanillas:** Conceptualization, Methodology,

<https://doi.org/10.1016/j.jclinepi.2022.04.027>

0895-4356/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Supervision, Project administration. **Rodrigo Fernández-López:** Investigation, Data Curation, Writing – Review & Editing. **Juan F. Huete:** Software, Validation, Resources, Writing – Review & Editing. **Carmen Amezcua-Prieto:** Validation, Visualization, Writing – Review & Editing. **Javier Zamora:** Methodology, Formal analysis, Resources, Writing – Review & Editing. **Juan M. Fernández-Luna:** Conceptualization, Methodology, Software, Validation, Writing – Review & Editing, Supervision. All authors contributed to the editing of the paper and approved its final version to be submitted.

* Corresponding author. Department of Preventive Medicine and Public Health, University of Granada. Avda. de la Investigación, 11. Granada 18016, Spain. Tel.: +34 958243544; fax: +34 958246118.

E-mail address: jrterceroh@gmail.com (J.R. Tercero-Hidalgo).

What is new?**Key findings**

- The use of artificial intelligence (AI) in COVID-19 systematic reviews was very low.
- COVID-19 reviews using AI tools showed higher publication impact and workload savings.

What this adds to what was known?

- Semi-automated screening and RCT filtering are the most notable use-cases of AI tools in evidence synthesis.
- There is a lack of systematic review tools cohesively integrating AI.

What is the implication and what should change now?

- There is scope for the application of AI in automating systematic reviews going forward.

1. Introduction

Evidence-based medicine depends on the production of timely systematic reviews to guide and update health care practice and policies [1]. This is a resource-intensive undertaking, requiring teams of multiple reviewers to interrogate numerous repositories and databases, screen through thousands of potentially relevant citations and articles, extract the pertinent data from the selected studies, and then prepare cohesive summaries of the findings [2,3]. In the context of the SARS-CoV2/COVID-19 pandemic, methods to speed up this lengthy process were urgently needed [4,5].

Systematic evidence synthesis relies on robust and standardized procedures to achieve dependable results. However, the call to accelerate research output during the pandemic led to a decrease on reviews' methodological quality [6,7] and the ascend of "rapid reviews" [8,9] (which shorten the usual timeframes by sacrificing on search depth, screening robustness or data extraction and at the expense of increased risk of errors). Are these unavoidable tradeoffs for timelier results?

Instead, artificial intelligence (AI) based solutions (that automate parts of the workflow by mimicking human problem-solving, comprising machine-learning, nature language processing, data mining and other subfields) [10] are now available to either complement or substitute human efforts with limited risk of bias [11–13], and have been previously (but scarcely) [14] employed in evidence synthesis to enhance screening [15] and data extraction [16,17]. Their aims are to shorten production times, allow for broader screenings of the literature and reduce reviewers' workloads without compromising on methodological quality.

Here, we evaluated the use of AI techniques among COVID-19 evidence syntheses to empirically determine whether, compared to COVID-19 evidence syntheses without AI, they impacted on the production, the quality, and the publication of systematic reviews.

2. Materials and methods

This methodological study [18] is reported following PRISMA 2020 guidelines [19] (checklist provided as [Supplementary material 1A](#)), and its protocol was prospectively registered at Open Science Forum Registries (DOI [10.17605/OSF.IO/H5DAW](#)) [20].

2.1. Search and selection of reviews

We considered for inclusion all COVID-19 related systematic reviews that could have made use of any AI tool (machine learning, deep learning, or natural language processing) to accelerate, improve or complement any aspect of the review conduct (search, screening, data extraction and synthesis). We implemented a script (available at DOI [10.5061/dryad.9kd51c5j6](#)) [21] to process all COVID-19 bibliographic references registered in the COVID-19 Living Overview of Evidence (L·OVE) database [22], filtering articles classified as "systematic review" between December 1st, 2019 and August 15th, 2021, and then querying the "Unpaywall" database [23] for every extracted DOI to obtain a JSON record with download links. The process was repeated three times since the publication of our protocol to reduce the loss of articles due to server-side errors (last searched on August 17th, 2021).

To capture reviews which deployed AI, we constructed a list of keywords with high probability of appearing in papers with AI tools ([Supplementary Material 1B](#)). We indexed every downloaded file with the OpenSemanticSearch search engine, running on a local Linux virtual machine. Every file that matched any of our keywords was manually inspected independently by two authors (JRTH and RFL). Pre-prints and non-English articles were included. The only exclusion criterion applied was non-open access status, due to the need to evaluate the methods section of each included review. To create a comparison group with sufficient statistical power of reviews without AI, for each included review we used the obtained records to randomly select three controls with the same publication date (within a 1-day margin if not enough articles were available for a given date). In addition, we located and included for analysis all previous versions of reviews labeled as living or "updated".

2.2. Data extraction

The following data were manually extracted independently by two authors (JRTH and RFL) from each review: type of review (as described by its authors: standard, rapid/scoping, living, or update of a prior version); disclosed

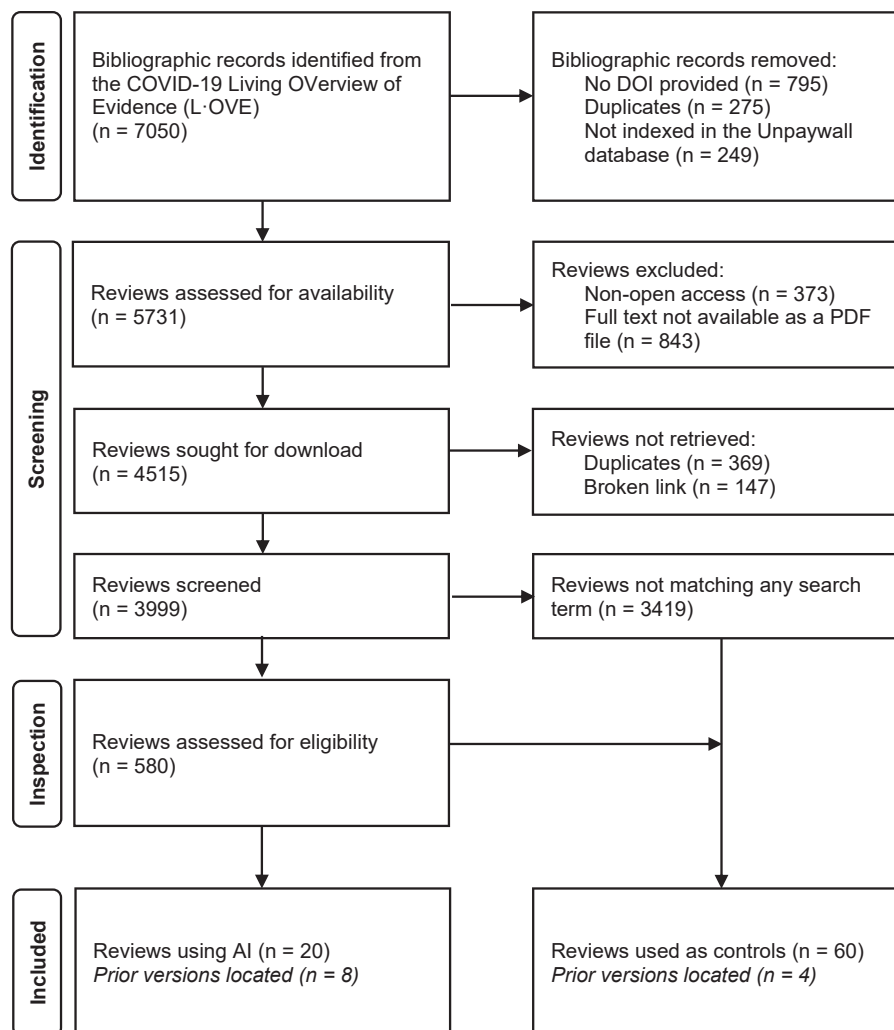


Fig. 1. Flowchart of included reviews: Flowchart of records obtained, screened, assessed for eligibility, and included in our study.

funding and conflicts of interest information; publication status, 2020 Journal Citation Reports (JCR) Impact Factor of the publishing journal and number of citations received (up to August 17th, 2021); number of abstracts screened, full texts reviewed and included studies; number of authors and of reviewers participating in the screening; and dates of protocol registration (if available) and of the review's earliest version. For living and updated reviews, we computed the increase in records screened and included between each of their versions and attributed their citation count to the newest one (to avoid double counting). Excel was used to record all variables.

Three authors (JRTH and RFL, assisted by CAP) graded all reviews with the AMSTAR-2 quality appraisal and risk of bias rating [24]. We excluded items 11-12 and 15, which apply to meta-analyses (as pre-specified by our protocol) and gave 0.5 points for "partial YES" answers when applicable, making for a maximum score of 13 points. For living and updated reviews, we only evaluated their most recent version (to avoid double counting). For reviews that

included both randomized controlled trials and observational studies, question 9 (assessment of the risk of bias of individual studies) was graded separately for each study type. The list of the quality items evaluated is provided as [Supplementary material IC](#).

2.3. Data synthesis

We calculated the ratios of abstracts screened and full texts inspected per author (as workload measurement) and per included study (screening precision). The number of reviewers participating in the screening was reported inconsistently between studies and was therefore not used in the calculations. We calculated the completion time of the pre-registered reviews as the difference between their protocol's date and the first pre-print's date of publication (or reception date at the journal, for published articles with no pre-prints available). Living and updated reviews' completion times were calculated as the difference between the publication dates of each of their versions. We excluded non pre-

Table 1. Extracted variables for artificial intelligence (AI) and control reviews: We used Pearson’s chi-square test to compare the proportions of rapid, living, funded, and published reviews, and the Wilcoxon–Mann–Whitney test for the rest of the comparisons. Medians and IQR (Q1–Q3) are rounded to the nearest integer

Characteristics	AI group (n = 20)		Controls (n = 60)		Δ	χ ²	P-value
	n	(%)	n	(%)			
Rapid reviews	5	(25%)	6	(10%)	15%	2.846	0.092
Living reviews	5	(25%)	3	(5%)	20%	6.667	0.010
Received funding	12	(60%)	27	(45%)	15%	1.351	0.245
Published	12	(60%)	48	(80%)	–20%	3.200	0.074
	Median	IQR	Median	IQR	Wilcoxon W		P-value
Journals’ JCR Impact Factor	9	(4–40)	3	(3–6)	409.0		<0.001
Citations per month	1	(0–13)	1	(0–3)	647.0		0.600
Abstracts screened							
Per author	302	(127–804)	140	(44–378)	1,126.0		0.009
Per included study	189	(94–366)	27	(14–64)	1,443.0		<0.001
Full texts inspected							
Per author	5	(4–16)	14	(7–37)	504.5		0.005
Per included study	4	(2–5)	3	(2–6)	883.5		0.481
Days to completion	74	(48–118)	123	(53–221)	183.5		0.205
AMSTAR-2 rating	8	(5–9)	6	(4–8)	740.5		0.119

Δ, absolute differences in percentage points between AI and control reviews; χ², test statistic for Pearson’s chi-square test; Wilcoxon W, test statistic for the Wilcoxon–Mann–Whitney rank sum test.

registered reviews from this metric due to heterogeneity in the reporting of their starting dates. We used Pearson’s chi-square test to compare the percentage of rapid, living, funded, and published reviews between groups. Publishing journals’ JCR Impact Factor, citation counts, screening workloads, completion times and AMSTAR-2 ratings were presented as medians with interquartile ranges (IQR), represented using box-and-whisker diagrams and compared using the Wilcoxon–Mann–Whitney test. R version 4.0.5 was used for statistical computing, and GraphPad Prism 9.2.0 for graphing. We also provided a narrative description of reviews using artificial intelligence, detailing which parts of the review process were automated and what software they used, how the AMSTAR-2 ratings differed among them, and how authors justified or what impact they attributed to the use of AI tools.

3. Results

3.1. Search and selection of reviews

As outlined in Figure 1, we identified 7,050 bibliographic records of COVID-19 systematic reviews, successfully downloaded 3,999, and manually inspected 580 that matched some of our keywords. We selected 20 reviews, of which there were 8 prior versions, making a total of 28 reviews (0.7% of the total, 95% CI 0.47–1.03%) with use of AI. Of the 60 articles selected as publication-date-matched controls, we located another 4 prior versions, making a total of 64 articles without use of AI. The complete list of selected articles is provided as an Excel document

(Supplementary Material 2, sheet “Included reviews”) with all the extracted variables and the AMSTAR-2 quality appraisal’s breakdown for each question. The full list of manually inspected and finally discarded articles is also provided (sheet “Excluded reviews”).

3.2. Description of the included reviews

Extracted variables are summarized in Table 1 and can be visualized in Figure 2. Of the 20 reviews selected for using AI, there were five rapid reviews (25%, with one scoping review and one rapid evidence map) and five living reviews (25%). Fifteen reviews provided a conflicts of interest statement, of which 12 (60%) declared having received external funding; 12 (60%) were published. Of the 60 control reviews, there were 6 rapid reviews (10%, with one scoping review) and three living reviews (5%). Fifty-seven reviews provided a conflicts of interest statement, of which 27 (45%) declared having received external funding; 48 (80%) were published. JCR Impact Factors and citation counts showed high variability in the AI group, mainly due to the inclusion of three BMJ [25–27], two Cochrane [28,29] and one Lancet [30] reviews. Furthermore, only 10 reviews in the AI group (50%) and 22 in the controls (36%) pre-registered a protocol, making for a total of 44 data points for the completion times’ calculation.

3.3. Comparison of AI reviews with controls

The AI group included a higher proportion of living reviews than the controls (5/20 vs. 3/60, 95% CI absolute

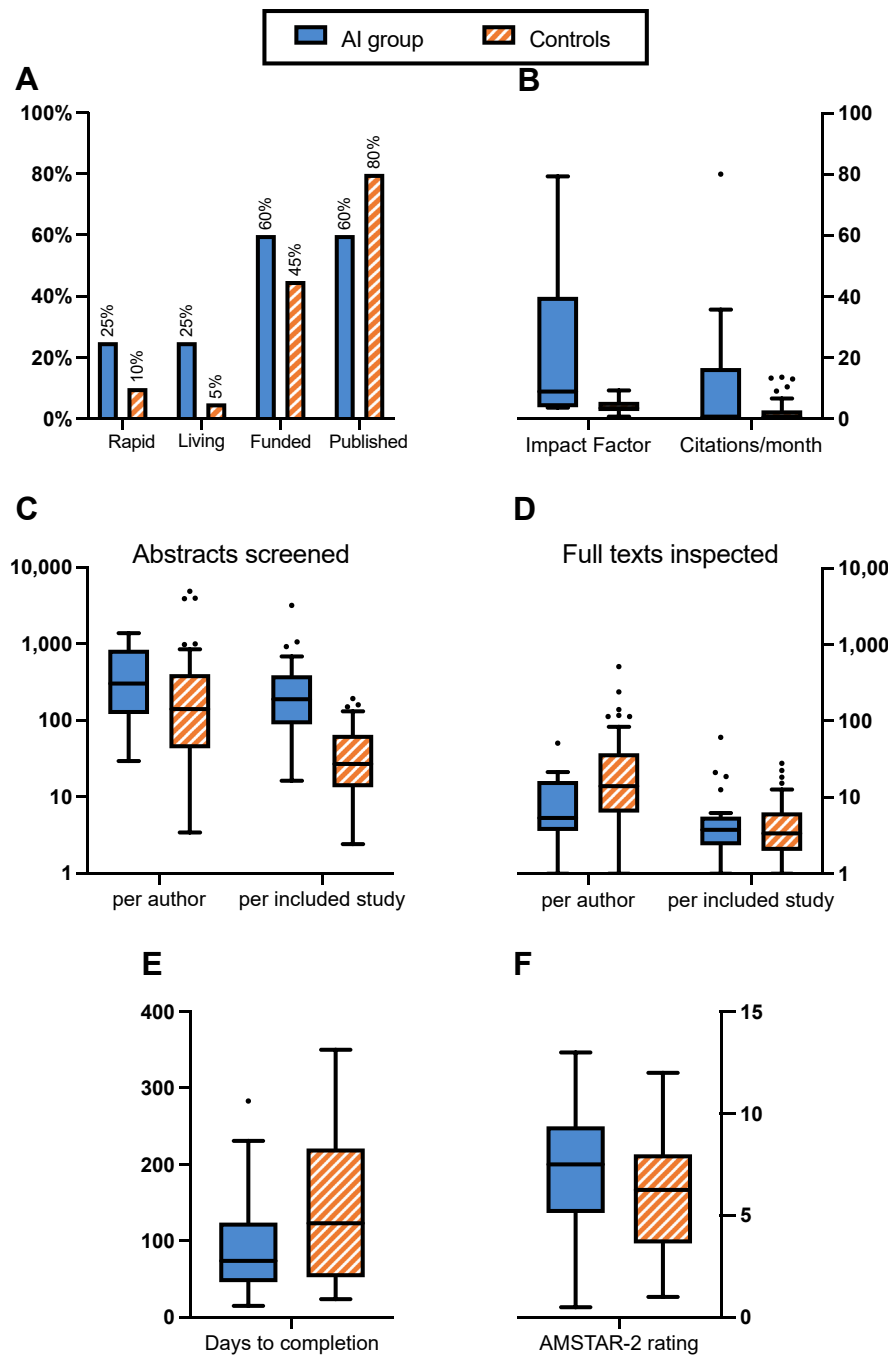


Fig. 2. Characteristics of the included reviews: Box-and-whisker diagram (the boxes enclose the Q1-Q3 quartiles, their middle lines represent the median, and whiskers extend to the furthest data points within 1.5 IQR). Panel A compares the proportion of rapid, living, funded, and published reviews between groups; Panel B presents the journals' 2020 JCR Impact Factors and citation counts of each group; Panels C and D show authors' workload measurements: abstracts screened and full-texts inspected, per author and per included study; Panel E exhibits the average times to completion (in days) of the reviews in each group; and Panel F represents their measured AMSTAR-2 ratings.

difference 0.2–39.8%, $P = 0.010$), while showing no differences in rapid reviews (5/20 vs. 6/60, 95% CI –5.4 to 35.4%, $P = 0.092$), funding (12/20 vs. 27/60, 95% CI –9.9 to 39.9%, $P = 0.245$) or publication status (12/20 vs. 48/60, 95% CI –43.7 to 3.7%, $P = 0.074$). JCR impact factors among published reviews in the AI group were significantly higher than the controls (median [IQR]: 8.9 [3.9–39.9] vs. 3.5 [2.6–5.5],

$P < 0.001$); citation counts showed no differences (0.5 [0.0–13.5] vs. 0.6 [0.0–2.8], $P = 0.600$).

Concerning the workload measurements, the AI group screened more abstracts per author (302.2 [126.7–804.3] vs. 140.3 [43.8–378.2], $P = 0.009$) and per included study (189.0 [94.1–365.8] vs. 26.9 [13.7–64.1], $P < 0.001$), while inspecting less full texts per author (5.3 [3.7–16.1]

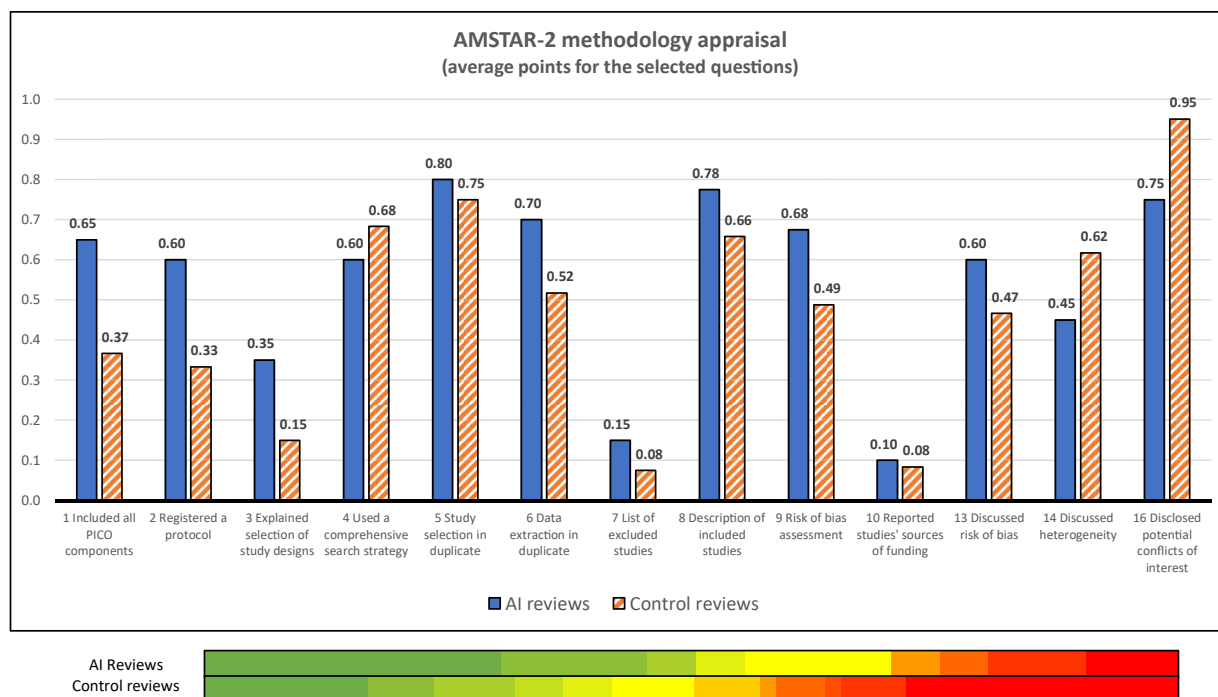


Fig. 3. AMSTAR-2 methodology appraisals' summary: The graph on the top shows the average ratings obtained in each of the evaluated questions by the reviews using Artificial Intelligence (AI) techniques (blue bars) and by the control group (orange bars). The colored bars on the bottom provide a visual representation of the quality appraisal's heterogeneity in both groups (a gradient was used to represent the obtained scores: red = 4; yellow = 6.5; green = 9).

vs. 14.0 [6.5–37.2], $P = 0.005$) and as many per included study (3.8 [2.4–5.3] vs. 3.4 [2.0–6.2], $P = 0.481$).

We observed no differences in the pre-registered reviews' times to completion (74.0 [47.5–117.5] vs. 123.0 [53.0–221.0], $P = 0.205$). The average scores obtained in the AMSTAR-2 risk of bias rating were not significantly higher in the AI group (7.5 [5.3–9.1] vs. 6.3 [3.9–8.0] points out of 13, $P = 0.119$), with both groups showing high heterogeneity of results as shown in Figure 3. Measured against the controls, the AI reviews scored worse on question 4 (literature search strategy, –12%) and better on question 6 (data extraction in duplicate, 35%), while showing minimal differences on question 5 (duplicate screening, 7%). Both groups scored the lowest on questions 7 (providing a list of excluded studies) and 10 (reporting on the sources of funding of the included studies).

3.4. Narrative description of the uses of AI in the included reviews

According to the step of the review process where AI was used, we can classify the 20 reviews in the AI group in three categories, as shown in Table 2.

3.4.1. Search process

Three reviews [31–33] complemented their search procedures with open-ended question queries on COVID-19 [45], an open dataset of COVID-19 related articles structured to facilitate the use of text mining and machine learning systems: Zaki et al. [32] used a GitHub repository

based on the Okapi BM25 search algorithm; Zaki et al. [33] employed BioBERT, a peer-reviewed [46] and open-source text mining system pre-trained for biomedical content analysis; and Parasa et al. [31] provided no details on the search engine employed. Additionally, Michelson et al. [34] used proprietary software from the “GenesisAI” company to produce a “rapid meta-analysis” as proof-of-concept of their product. Daley et al. [35] disclosed no information on the software employed. Only two reviews in this subgroup were published, and none registered a protocol. The average AMSTAR-2 score was 3.7/13.

3.4.2. Filtering of randomized controlled trials

Seven articles [25,26,36–40] employed RobotSearch, a peer-reviewed [47] and open-source software to identify randomized controlled trials (RCT) from a citations list. It is based on a neural network trained with data from Cochrane's reviews and stands out for its ease of use (no installation is required) and flexibility (as it allows for different levels of sensitivity, including one developed specifically for systematic reviews, as well as integration with other scripts).

In our sample, RobotSearch was often incorporated in the workflows of living or partially automated reviews. Two of the reviews that made use of RobotSearch were Bartoszko et al. [25], a network meta-analysis of the evidence for COVID-19 prophylaxis, and Siemieniuk et al. [26], a living meta-analysis of randomized trials to inform World Health Organization (WHO) Living Guidelines on

Table 2. AI tools used in COVID-19 reviews: Table showing the different artificial intelligence (AI) tools that have been used in the elaboration of COVID-19 systematic reviews, according to their area of application: search assistance, randomized controlled trials (RCT) filtering and screening automation

Ref.	Title	Authors	Journal	AI used in...	Software used	Is open source?
[31]	Prevalence of Gastrointestinal Symptoms and Fecal Viral Shedding in Patients with Coronavirus Disease 2019	Parasa et al.	JAMA Network Open	Search	CORD-19	Partially
[32]	The influence of comorbidity on the severity of COVID-19 disease: systematic review and analysis	Zaki et al.	Pre-print	Search	CORD-19 + Okapi BM25	Yes
[33]	The Estimations of the COVID-19 Incubation Period: A Scoping Reviews of the Literature	Zaki et al.	Journal of Infection and Public Health	Search	CORD-19 + BioBERT	Yes
[34]	Ocular toxicity and Hydroxychloroquine: A Rapid Meta-Analysis	Michelson et al.	Pre-print	Search	GenesisAI (formerly Evid Science)	No
[35]	A Systematic Review of the Incubation Period of SARS-CoV-2: The Effects of Age, Biological Sex, and Location on Incubation Period	Daley et al.	Pre-print	Search	Not reported	No
[36]	Impact of remdesivir on 28 day mortality in hospitalized patients with COVID-19: February 2021 Meta-analysis	Robinson et al.	Pre-print	RCT filtering	RobotSearch	Yes
[37]	Impact of systemic corticosteroids on hospitalized patients with COVID-19: January 2021 Meta-analysis of randomized controlled trials	Robinson et al.	Pre-print	RCT filtering	RobotSearch	Yes
[25]	Prophylaxis against COVID-19: living systematic review and network meta-analysis	Bartoszko et al.	BMJ	RCT filtering	RobotSearch	Yes
[26]	Drug treatments for COVID-19: living systematic review and network meta-analysis	Siemieniuk et al.	BMJ	RCT filtering	RobotSearch	Yes
[38]	Adverse effects of remdesivir, hydroxychloroquine, and lopinavir/ritonavir when used for COVID-19: systematic review and meta-analysis of randomized trials	Izcovich et al.	Pre-print	RCT filtering	RobotSearch	Yes
[39]	Tocilizumab and sarilumab alone or in combination with corticosteroids for COVID-19: A systematic review and network meta-analysis	Zeraatkar et al.	Pre-print	RCT filtering	RobotSearch	Yes

(Continued)

Table 2. Continued

Ref.	Title	Authors	Journal	AI used in...	Software used	Is open source?
[40]	Clinical trials in COVID-19 management & prevention: A meta-epidemiological study examining methodological quality	Honarmand et al.	Journal of Clinical Epidemiology	RCT filtering	RobotSearch	Yes
[41]	Impacts of school closures on physical and mental health of children and young people: a systematic review	Viner et al.	Pre-print	Screening	EPPI-Reviewer	No
[27]	Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal	Wynants et al.	BMJ	Screening	EPPI-Reviewer	No
[28]	Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection (Review)	Dinnes et al.	Cochrane Database of Systematic Reviews	Screening	EPPI-Reviewer	No
[29]	Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19	Struyf et al.	Cochrane Database of Systematic Reviews	Screening	EPPI-Reviewer	No
[42]	Are medical procedures that induce coughing or involve respiratory suctioning associated with increased generation of aerosols and risk of SARS-CoV-2 infection? A rapid systematic review	Wilson et al.	Journal of Hospital Infection	Screening	EPPI-Reviewer	No
[43]	Risk and Protective Factors in the COVID-19 Pandemic: A Rapid Evidence Map	Elmore et al.	Frontiers in Public Health	Screening	SWIFT-Active Screener	No
[44]	Tocilizumab and Systemic Corticosteroids in the Management of COVID-19 Patients: A Systematic Review and Meta-Analysis	Alkofide et al.	International Journal of Infectious Diseases	Screening	Abstrackr	Yes
[30]	Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis	Chu et al.	The Lancet	Screening	Evidence Prime	No

drugs for treatment of COVID-19, of which Izcovich et al. [38] and Zeraatkar et al. [39] are separate sub-studies. Both are part of the “*BMJ Rapid Recommendations*” project and maintain a website where summaries of the evidence available and interim analyses are published. The average AMSTAR-2 score was 7.5/13.

3.4.3. Screening of titles and abstracts

We found eight articles [27–30,41–44] that made use of AI-powered screening procedures. Five of them [27–29,41,42] used EPPI-Reviewer, a web-based tool (distributed as shareware) to assist in the elaboration of all kinds of literature reviews. It offers a wide variety of features, from bibliographic management to collaborative working, as well as study identification capabilities, automatic clustering of articles, and text mining. In particular, the included reviews used its “SGCClassifier” module to prioritize the screening of articles more likely to be included. As a result, both Wynants et al. [27] and two Cochrane reviews [28,29] quoted a 80% reduction in the screening burden due to this tool.

Similar screening automation techniques from systematic reviews’ elaboration platforms were used by other two articles: SWIFT-Active Screener [48] by Elmore et al. [43], which was set to achieve a certain study recall objective as the screening’s stopping criterion; and Evidence Prime by Chu et al. [30], to double-check the screening process. Finally, Alkofide et al. [44] used Abstrackr, the only open-source software in this category, which uses feedback from previously selected and rejected articles to guide the screening process. Evaluations of this tool published in the literature [49] suggest high workload savings in the production of systematic reviews at the cost of 0.1% false negative rates.

Among the reviews analyzed in this study, this subgroup presented the highest scores in the AMSTAR-2 appraisal tool (9.1/13), with the notable mentions of two Cochrane reviews [28,29] (12 points) and a rapid meta-analysis [30] published in the *Lancet* (10.5 points). Contrary to reviews in the other categories that prioritized search depth, the use of AI-powered tools in this subgroup was motivated by the screening burden faced by the reviewers: quoting Dinnes et al. [28], “*a more efficient approach [was needed] to keep up with the rapidly increasing volume of COVID-19 literature*”.

4. Discussion

We evaluated if the potential benefits of deploying AI in evidence syntheses have been realized in COVID-19 reviews. We found that AI was rarely utilized, appearing in only 0.7% of the studied reviews, but that it was significantly associated with reductions in authors’ screening workload and publication in journals with higher Impact Factor. Being a living review was associated with using AI, with the most common use cases being the

optimization of screening (prioritizing studies with high likelihood of being relevant) and the selection of randomized controlled trials.

As a limitation of our study, we would highlight its low statistical power due to the small number of reviews using AI. Anticipating the limited availability of reviews with AI, we adopted a highly sensitive screening procedure, processing more than 7,000 bibliographic references of COVID-19 systematic reviews (combining expert advice in the selection of keywords and a fully-featured search engine), and chose a 3:1 control group size to minimize the risk of type II statistical errors. Using L-OVE as our primary database allowed access to all relevant and updated sources in a systematic and machine-readable way; however, our search strategy might show a reduced sensitivity to institutional reports and white-papers, often not indexed by traditional databases. The impact of download errors and excluding non-open-access reviews from our study is uncertain; its influence on generalizing our results should be interpreted in light of the diversity of secondary sources reachable through L-OVE and the high accessibility of COVID-19 research during the pandemic. Furthermore, the use of publication dates as a matching variable allowed for a bias-minimizing (script-driven) selection of controls but it prevented the use of other desirable controlling variables such as review sizes or goals.

We also note that reporting workloads “per author” instead of “per reviewer participating in the screening” may underestimate workload measurements for large teams (when not all their authors participate in the screening). A higher author count might also be related to resource availability, and thus access to expert advice regarding AI. Likewise, better-resourced groups with AI expert support might have greater access to well-indexed journals, potentially biasing Impact Factor analyses in favor of AI. The AMSTAR-2 tool was inevitably applied without blinding the reviewers to use or non-use of AI, which, given the subjectiveness of certain aspects of the methodology assessment, might have influenced this evaluation. Finally, the use of citation counts to measure reviews’ impact has known deficiencies such as being influenced by citation bias or the authority of the authors [50], and this approach may underestimate the impact of recently published reports.

On average, it takes 15 months for teams of five reviewers to complete a traditional systematic review [51], with estimated screening error rates of around 10% [52]. Facing the COVID-19 pandemic demanded robust evidence summaries with urgency as delays incurred cost in terms of lost lives and economic damage. However, despite the explosive growth that the AI and machine learning fields have experienced during the last years, they played a surprisingly limited role in COVID-19 evidence synthesis. Our findings are consistent with previous reports [14] that the benefits AI can provide in the conduct of systematic reviews are unknown to most review authors, while the relative unorthodoxy of its methods might initially hinder their acceptance by the research community. Open-source software, more prone to community adoption, will be essential in this

aspect. Hopefully, our article will raise the profile of AI in evidence syntheses.

Our narrative description of the reviews included in this study showed that none made use of more than one AI-tool. A more cohesive approach, seamlessly merging AI into every step of the review process, would save reviewers' time trying to interconnect different tools with sometimes incompatible formats. Semi-automated screening procedures were one of the areas where AI showed more adoption, and the variety of software options (such as EPPI-Reviewer, already adopted as a Cochrane Review Production Tool) was higher. On the contrary, full automation was only employed by RobotSearch (an extensively appraised randomized trials identifier), suggesting that the adoption of increasingly automated solutions may be hindered by the need to further assess their potential cost on recall and risk-of-bias against their productivity contributions.

5. Conclusion

The need for automated solutions in research synthesis is obvious, as reviewers' workload is growing with the rapidly expanding biomedical field. Adoption of new technologies can take time, but realizing AI's potential in evidence synthesis should be a priority. Going forward, AI must be incorporated to systematic reviews as the next step toward timely, better, and more responsive decision-making.

CRedit authorship contribution statement

Juan R. Tercero-Hidalgo: Methodology, Software, Investigation, Writing – Original Draft. **Khalid S. Khan:** Conceptualization, Methodology, Writing – Review & Editing, Supervision. **Aurora Bueno-Cavanillas:** Conceptualization, Methodology, Supervision, Project administration. **Rodrigo Fernández-López:** Investigation, Data Curation, Writing – Review & Editing. **Juan F. Huete:** Software, Validation, Resources, Writing – Review & Editing. **Carmen Amezcua-Prieto:** Validation, Visualization, Writing – Review & Editing. **Javier Zamora:** Methodology, Formal analysis, Resources, Writing – Review & Editing. **Juan M. Fernández-Luna:** Conceptualization, Methodology, Software, Validation, Writing – Review & Editing, Supervision. All authors contributed to the editing of the paper and approved its final version to be submitted.

Appendix A

Supplementary Data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2022.04.027>.

References

- [1] Lasserson TJ, Thomas J, Higgins JPT. Chapter 1: Starting a review. Cochrane handbook for systematic reviews of interventions. Cochrane; 2022: 6.3 (updated February 2022). Available at: www.training.cochrane.org/handbook.
- [2] Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. *Syst Rev* 2014;3:1–15. <https://doi.org/10.1186/2046-4053-3-74>.
- [3] Nussbaumer-Streit B, Ellen M, Klerings I, Sftcu R, Riva N, Mahmić-Kaknjo M, et al. Resource use during systematic review production varies widely: a scoping review. *J Clin Epidemiol* 2021;139:287–96.
- [4] Gill D, Baker EH, Hitchings AW. We need clinical guidelines fit for a pandemic. *BMJ* 2021;373:n1093. <https://doi.org/10.1136/bmj.n1093>.
- [5] Knottnerus JA, Tugwell P. Methodological challenges in studying the COVID-19 pandemic crisis. *J Clin Epidemiol* 2020;121:A5–7.
- [6] Li Y, Cao L, Zhang Z, Hou L, Qin Y, Hui X, et al. Reporting and methodological quality of COVID-19 systematic reviews needs to be improved: an evidence mapping. *J Clin Epidemiol* 2021;135:17–28.
- [7] Jung RG, Di Santo P, Clifford C, Prosperi-Porta G, Skanes S, Hung A, et al. Methodological quality of COVID-19 clinical research. *Nat Commun* 2021;12:1–10. <https://doi.org/10.1038/s41467-021-21220-5>.
- [8] Tricco AC, Garrity CM, Boulos L, Lockwood C, Wilson M, McGowan J, et al. Rapid review methods more challenging during COVID-19: commentary with a focus on 8 knowledge synthesis steps. *J Clin Epidemiol* 2020;126:177–83.
- [9] Biesty L, Meskell P, Glenton C, Delaney H, Smalle M, Booth A, et al. A QuESr for speed: rapid qualitative evidence syntheses as a response to the COVID-19 pandemic. *Syst Rev* 2020;9:1–6. <https://doi.org/10.1186/s13643-020-01512-5>.
- [10] Amezcua-Prieto C, Fernández-Luna JM, Huete-Guadix JF, Bueno-Cavanillas A, Khan KS. Artificial intelligence and automation of systematic reviews in women's health. *Curr Opin Obstet Gynecol* 2020;32:335–41.
- [11] O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Syst Rev* 2015;4:1–22. <https://doi.org/10.1186/2046-4053-4-5>.
- [12] Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Syst Rev* 2019;8:1–10. <https://doi.org/10.1186/s13643-019-1074-9>.
- [13] Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, et al. Living systematic reviews: 2. Combining human and machine effort. *J Clin Epidemiol* 2017;91:31–7.
- [14] Scott AM, Forbes C, Clark J, Carter M, Glasziou P, Munn Z. Systematic review automation tools improve efficiency but lack of knowledge impedes their adoption: a survey. *J Clin Epidemiol* 2021;138:80–94.
- [15] Thomas J, McDonald S, Noel-Storr A, Shemilt I, Elliott J, Mavergames C, et al. Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews. *J Clin Epidemiol* 2021;133:140–51.
- [16] Schmidt L, Olorisade BK, McGuinness LA, Thomas J, Higgins JPT. Data extraction methods for systematic review (semi)automation: a living systematic review [version 1; peer review: 3 approved]. *F1000Res* 2021;10:1–35. <https://doi.org/10.12688/f1000research.51117.1>.
- [17] Jonnalagadda SR, Goyal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. *Syst Rev* 2015;4:1–16. <https://doi.org/10.1186/s13643-015-0066-7>.
- [18] Mbuagbaw L, Lawson DO, Puljak L, Allison DB, Thabane L. A tutorial on methodological studies: the what, when, how and why. *BMC*

- 134 *J.R. Tercero-Hidalgo et al. / Journal of Clinical Epidemiology 148 (2022) 124–134*
- Med Res Methodol 2020;20:1–12. <https://doi.org/10.1186/s12874-020-01107-7>. 226.
- [19] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
- [20] Tercero-Hidalgo JR, Khan KS, Bueno-Cavanillas A, Fernández-López R, Huete JF, Amezcua-Prieto C, et al. Covid-19 systematic evidence synthesis with artificial intelligence: a review of reviews. *Open Sci Forum Regist* 2021. <https://doi.org/10.17605/OSF.IO/H5DAW>.
- [21] Tercero-Hidalgo JR, Khan KS, Bueno-Cavanillas A, Fernández-López R, Huete JF, Amezcua-Prieto C, et al. COVID-19 evidence syntheses with artificial intelligence: an empirical study of systematic reviews. *Dryad Dataset* 2021. <https://doi.org/10.5061/dryad.9kd51c5j6>.
- [22] Rada G, Verdugo-Paiva F, Ávila C, Morel-Marambio M, Bravo-Jeria R, Pesce F, et al. Evidence synthesis relevant to COVID-19: a protocol for multiple systematic reviews and overviews of systematic reviews. *Medwave* 2020;20:e7868.
- [23] Dhakal K. *Unpaywall*. *J Med Libr Assoc* 2019;107:286–8.
- [24] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. Amstar 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008.
- [25] Bartoszko JJ, Siemieniuk RAC, Kum E, Qasim A, Zeraatkar D, Ge L, et al. Prophylaxis against covid-19: living systematic review and network meta-analysis. *BMJ* 2021;373:n949.
- [26] Siemieniuk RAC, Bartoszko JJ, Ge L, Zeraatkar D, Izcovich A, Kum E, et al. Drug treatments for covid-19: living systematic review and network meta-analysis. *BMJ* 2020;370:m2980.
- [27] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020;369:m1328.
- [28] Dinnes J, Deeks JJ, Berhane S, Taylor M, Adriano A, Davenport C, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst Rev* 2021; 3:CD013705.
- [29] Struyf T, Deeks JJ, Dinnes J, Takwoingi Y, Davenport C, Leeftang MMG, et al. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database Syst Rev* 2021;2:CD013665.
- [30] Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *Lancet* 2020;395:1973–87.
- [31] Parasa S, Desai M, Thogulva Chandrasekar V, Patel HK, Kennedy KF, Roesch T, et al. Prevalence of gastrointestinal symptoms and fecal viral shedding in patients with coronavirus disease 2019: a systematic review and meta-analysis. *JAMA Netw Open* 2020;3:e2011335.
- [32] Zaki N, Mohamed EA, Ibrahim S, Khan G. The influence of comorbidity on the severity of COVID-19 disease: a systematic review and analysis. *medRxiv* 20201–17. <https://doi.org/10.1101/2020.06.18.20134478>. 2020.06.18.20134478.
- [33] Zaki N, Mohamed EA. The estimations of the COVID-19 incubation period: a scoping reviews of the literature. *J Infect Public Health* 2021;14:638–46.
- [34] Michelson M, Chow T, Martin N, Ross M, Tee A, Minton S. Ocular toxicity and hydroxychloroquine: a rapid meta-analysis. *MedRxiv* 2020;22:e20007.
- [35] Daley C, Fydenkevez M, Ackerman-Morris S. A systematic review of the incubation period of SARS-CoV-2: the effects of age, biological sex, and location on incubation period. *MedRxiv* 20201–19. <https://doi.org/10.1101/2020.12.23.20248790>. 2020.12.23.20248790.
- [36] Robinson R, Prakash V, Tamimi R Al, Albast N, Al-Bast B, Wieland E, et al. Impact of remdesivir on 28 day mortality in hospitalized patients with COVID-19: February 2021 Meta-analysis. *MedRxiv* 20211–15. <https://doi.org/10.1101/2021.03.04.21252903>. 2021.03.04.21252903.
- [37] Robinson R, Prakash V, Tamimi R Al, Albast N, Al-Bast B. Impact of systemic corticosteroids on hospitalized patients with COVID-19: January 2021 Meta-analysis of randomized controlled trials. *MedRxiv* 20211–15. <https://doi.org/10.1101/2021.02.03.21251065>. *medRxiv* 2021.02.03.21251065.
- [38] Izcovich A, Siemieniuk RAC, Bartoszko JJ, Ge L, Zeraatkar D, Kum E, et al. Adverse effects of remdesivir, hydroxychloroquine, and lopinavir/ritonavir when used for COVID-19: systematic review and meta-analysis of randomized trials. *BMJ Open* 2022;12(3):1–12. <https://doi.org/10.1136/bmjopen-2020-048502>. e048502.
- [39] Zeraatkar D, Cusano E, Martínez JPD, Qasim A, Mangala SO, Kum E, et al. Use of tocilizumab and sarilumab alone or in combination with corticosteroids for covid-19: systematic review and network meta-analysis. *BMJ Medicine* 2022;1(1):1–14. <https://doi.org/10.1101/2021.07.05.21259867>. e000036.
- [40] Honarmand K, Penn J, Agarwal A, Siemieniuk R, Brignardello-Petersen R, Bartoszko JJ, et al. Clinical trials in COVID-19 management & prevention: a meta-epidemiological study examining methodological quality. *J Clin Epidemiol* 2021;139:68–79.
- [41] Viner R, Russell S, Saull R, Croker H, Stansfeld C, Packer J, et al. Impacts of school closures on physical and mental health of children and young people: a systematic review. *MedRxiv* 20201–31. <https://doi.org/10.1101/2021.02.10.21251526>. 2021.02.10.21251526.
- [42] Wilson J, Carson G, Fitzgerald S, Llewelyn MJ, Jenkins D, Parker S, et al. Are medical procedures that induce coughing or involve respiratory suctioning associated with increased generation of aerosols and risk of SARS-CoV-2 infection? A rapid systematic review. *J Hosp Infect* 2021;116:37–46.
- [43] Elmore R, Schmidt L, Lam J, Howard BE, Tandon A, Norman C, et al. Risk and protective factors in the COVID-19 pandemic: a rapid evidence map. *Front Public Heal* 2020;8:582205.
- [44] Alkofide H, Almohaizeie A, Almuhaimee S, Alotaibi B, Alkharfy KM. Tocilizumab and systemic corticosteroids in the management of patients with COVID-19: a systematic review and meta-analysis. *Int J Infect Dis* 2021;110:320–9.
- [45] Lu Wang L, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, et al. COVID-19: the COVID-19 Open Research Dataset. *ArXiv* 20201–11. <https://doi.org/10.48550/arXiv.2004.10706>. 2004.10706v4.
- [46] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020;36:1234–40.
- [47] Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. Machine learning for identifying randomized controlled trials: an evaluation and practitioner’s guide. *Res Synth Methods* 2018;9:602–14.
- [48] Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, et al. SWIFT-Active Screener: accelerated document screening through active learning and integrated recall estimation. *Environ Int* 2020;138:105623.
- [49] Gates A, Johnson C, Hartling L. Technology-assisted title and abstract screening for systematic reviews: a retrospective evaluation of the Abstrackr machine learning tool. *Syst Rev* 2018;7:1–9. <https://doi.org/10.1186/s13643-018-0707-8>. 45.
- [50] Urralings MJE, Duyx B, Swaen GMH, Bouter LM, Zeegers MP. Citation bias and other determinants of citation in biomedical research: findings from six citation networks. *J Clin Epidemiol* 2021;132: 71–8.
- [51] Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open* 2017;7:e012545.
- [52] Wang Z, Nayfeh T, Tetzlaff J, O’Blenis P, Murad MH. Error rates of human reviewers during abstract screening in systematic reviews. *PLoS One* 2020;15:e0227742.

Apéndice B

Checklist PRISMA 2020

Tabla B.1 Checklist PRISMA 2020

Sección/tema	Nº	Ítem de la lista de verificación	Localización
TÍTULO	1	Identifique la publicación como una revisión sistemática.	1
RESUMEN	2	Vea la lista de verificación para resúmenes estructurados de la declaración PRISMA 2020.	Tabla B.2
INTRODUCCIÓN			
Justificación	3	Describa la justificación de la revisión en el contexto del conocimiento existente.	13
Objetivos	4	Proporcione una declaración explícita de los objetivos o las preguntas que aborda la revisión.	14
MÉTODOS			
Criterios de elegibilidad	5	Especifique los criterios de inclusión y exclusión de la revisión y cómo se agruparon los estudios para la síntesis.	15
Fuentes de información	6	Especifique todas las bases de datos, registros, sitios web, organizaciones, listas de referencias y otros recursos de búsqueda o consulta para identificar los estudios. Especifique la fecha en la que cada recurso se buscó o consultó.	14-15
Estrategia de búsqueda	7	Presente las estrategias de búsqueda completas de todas las bases de datos, registros y sitios web, incluyendo cualquier filtro y los límites utilizados.	15
Proceso de selección de los estudios	8	Especifique los métodos utilizados para decidir si un estudio cumple con los criterios de inclusión de la revisión, incluyendo cuántos autores de la revisión cribaron cada registro y cada publicación recuperada, si trabajaron de manera independiente y, si procede, los detalles de las herramientas de automatización utilizadas en el proceso.	15
Proceso de extracción de los datos	9	Indique los métodos utilizados para extraer los datos de los informes o publicaciones, incluyendo cuántos revisores recopilaron datos de cada publicación, si trabajaron de manera independiente, los procesos para obtener o confirmar los datos por parte de los investigadores del estudio y, si procede, los detalles de las herramientas de automatización utilizadas en el proceso.	15
Lista de los datos	10a	Enumere y defina todos los desenlaces para los que se buscaron los datos. Especifique si se buscaron todos los resultados compatibles con cada dominio del desenlace (por ejemplo, para todas las mediciones, puntos temporales, análisis) y, de no ser así, los métodos utilizados para decidir los resultados que se debían recoger.	15-16
	10b	Enumere y defina todas las demás variables para las que se buscaron datos (por ejemplo, características de los participantes y de la intervención, fuentes de financiación). Describa todos los supuestos formulados sobre cualquier información ausente (missing) o incierta.	16-17

Tabla B.1 Checklist PRISMA 2020 (continuación)

Sección/tema	Nº	Ítem de la lista de verificación	Localización
Evaluación del riesgo de sesgo de los estudios individuales	11	Especifique los métodos utilizados para evaluar el riesgo de sesgo de los estudios incluidos, incluyendo detalles de las herramientas utilizadas, cuántos autores de la revisión evaluaron cada estudio y si trabajaron de manera independiente y, si procede, los detalles de las herramientas de automatización utilizadas en el proceso.	16
Medidas del efecto	12	Especifique, para cada desenlace, las medidas del efecto (por ejemplo, razón de riesgos, diferencia de medias) utilizadas en la síntesis o presentación de los resultados.	16-17
Métodos de síntesis	13a	Describa el proceso utilizado para decidir qué estudios eran elegibles para cada síntesis (por ejemplo, tabulando las características de los estudios de intervención y comparándolas con los grupos previstos para cada síntesis (ítem n.º 5)).	No aplicable
	13b	Describa cualquier método requerido para preparar los datos para su presentación o síntesis, tales como el manejo de los datos faltantes en los estadísticos de resumen o las conversiones de datos.	17
	13c	Describa los métodos utilizados para presentar visualmente los resultados de los estudios individuales y su síntesis.	17
	13d	Describa los métodos utilizados para sintetizar los resultados y justifique sus elecciones. Si se ha realizado un metanálisis, describa los modelos, los métodos para identificar la presencia y el alcance de la heterogeneidad estadística, y los programas informáticos utilizados.	17
	13e	Describa los métodos utilizados para explorar las posibles causas de heterogeneidad entre los resultados de los estudios (por ejemplo, análisis de subgrupos, metarregresión).	17
	13f	Describa los análisis de sensibilidad que se hayan realizado para evaluar la robustez de los resultados de la síntesis.	No aplicable
Evaluación del sesgo en el informe	14	Describa los métodos utilizados para evaluar el riesgo de sesgo debido a resultados faltantes en una síntesis (derivados de los sesgos en el informe).	No aplicable
Evaluación de la certeza	15	Describa los métodos utilizados para evaluar la certeza (o confianza) en el cuerpo de la evidencia para cada desenlace.	No aplicable
RESULTADOS			
Selección de los estudios	16a	Describa los resultados de los procesos de búsqueda y selección, desde el número de registros identificados en la búsqueda hasta el número de estudios incluidos en la revisión, idealmente utilizando un diagrama de flujo (Figura 1).	18-19
	16b	Cite los estudios que aparentemente cumplían con los criterios de inclusión, pero que fueron excluidos, y explique por qué fueron excluidos.	(*)
Características	17	Cite cada estudio incluido y presente sus características.	24-25
Riesgo de sesgo de los estudios individuales	18	Presente las evaluaciones del riesgo de sesgo para cada uno de los estudios incluidos.	(*)

Tabla B.1 Checklist PRISMA 2020 (continuación)

Sección/tema	Nº	Ítem de la lista de verificación	Localización
Resultados de los estudios individuales	19	Presente, para todos los desenlaces y estudios: a) los estadísticos de resumen para cada grupo (si procede) y b) la estimación del efecto y su precisión (p.ej., intervalo de confianza), idealmente con tablas estructuradas o gráficos.	No aplicable
Resultados de la síntesis	20a	Para cada síntesis, resuma brevemente las características y el riesgo de sesgo entre los estudios contribuyentes.	19-20
	20b	Presente los resultados de todas las síntesis estadísticas realizadas. Si se ha realizado un metanálisis, presente para cada uno de ellos el estimador de resumen y su precisión (por ejemplo, intervalo de credibilidad o de confianza) y las medidas de heterogeneidad estadística. Si se comparan grupos, describa la dirección del efecto.	20-23
	20c	Presente las investigaciones sobre las posibles causas de heterogeneidad entre los resultados de los estudios.	26-28
	20d	Presente los resultados de todos los análisis de sensibilidad realizados para evaluar la robustez de los resultados.	No aplicable
Sesgos en el informe	21	Presente las evaluaciones del riesgo de sesgo debido a resultados faltantes para cada síntesis evaluada.	No aplicable
Certeza de la evidencia	22	Presente las evaluaciones de la certeza (o confianza) en el cuerpo de la evidencia para cada desenlace evaluado.	No aplicable
DISCUSIÓN	23a	Proporcione una interpretación general de los resultados en el contexto de otras evidencias.	28
	23b	Argumente las limitaciones de la evidencia incluida en la revisión.	29
	23c	Argumente las limitaciones de los procesos de revisión utilizados.	29-30
	23d	Argumente las implicaciones de los resultados para la práctica, las políticas y las futuras investigaciones.	31
OTRA INFORMACIÓN			
Registro y protocolo	24a	Proporcione la información del registro de la revisión, o declare que la revisión no ha sido registrada.	6, 14
	24b	Indique dónde se puede acceder al protocolo, o declare que no se ha redactado ningún protocolo.	6, 14
	24c	Describa y explique cualquier enmienda a la información proporcionada en el registro o en el protocolo.	16-17
Financiación	25	Describa las fuentes de apoyo financiero o no financiero para la revisión y el papel de los financiadores o patrocinadores..	6
Conflicto de intereses	26	Declare los conflictos de intereses de los autores de la revisión.	6
Disponibilidad de datos, códigos y otros materiales	27	Especifique qué elementos de los que se indican a continuación están disponibles al público y dónde se pueden encontrar: plantillas de formularios de extracción de datos, datos extraídos de los estudios incluidos, datos utilizados para todos los análisis, código de análisis, cualquier otro material utilizado en la revisión.	6, 14, 18

Adaptado de: [19] Page MJ et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. Traducción disponible en: *Rev Española Cardiol*. 2021;74:790-9.

(*) *Suppl.mat. 2: disponible en la web de la revista.*

Tabla B.2 Checklist para Abstracts PRISMA 2020

Sección/tema	Nº	Ítem de la lista de verificación	¿Presente?
TÍTULO			
Título	1	Identifique el informe como una revisión sistemática.	Sí
ANTECEDENTES			
Objetivos	3	Proporcionar una declaración explícita de los principales objetivos o preguntas que aborda la revisión.	Sí
MÉTODOS			
Criterios de elegibilidad	4	Especifique los criterios de inclusión y exclusión de la revisión.	Sí
Fuentes de información	5	Especifique las fuentes de información (por ejemplo, bases de datos, registros) utilizadas para identificar los estudios y la fecha de la última búsqueda en cada una de ellas.	Sí
Riesgo de sesgo	6	Especifique los métodos utilizados para evaluar el riesgo de sesgo en los estudios incluidos.	Sí
Síntesis de los resultados	6	Especifique los métodos utilizados para presentar y sintetizar los resultados.	No
RESULTADOS			
Estudios incluidos	6	Indicar el número total de estudios y participantes incluidos y resumir las características relevantes de los estudios.	Sí
Síntesis de los resultados	6	Presentar los resultados de los principales resultados, indicando preferentemente el número de estudios incluidos y de participantes para cada uno. Si se realizó un meta-análisis, informe la estimación resumida y el intervalo de confianza/creencia. Si se comparan grupos, indicar la dirección del efecto (es decir, qué grupo se ve favorecido).	Sí
DISCUSIÓN			
Limitaciones de la evidencia	6	Proporcione un breve resumen de las limitaciones de la evidencia incluida en la revisión (por ejemplo, riesgo de sesgo del estudio, inconsistencia e imprecisión).	No
Interpretación	6	Proporcionar una interpretación general de los resultados y las implicaciones importantes.	Sí
OTROS			
Financiación	6	Especifique la fuente principal de financiación de la revisión.	No
Registro	6	Proporcionar el nombre del registro y el número de registro.	No

Apéndice C

Palabras clave de búsqueda

- | | | |
|-------------------------------------|----------------------------------|---------------------------------|
| 1. support vector machine | 32. maximum entropy classifier | 63. machine learning |
| 2. SVM | 33. embedding | 64. speech recognition |
| 3. logistic regression | 34. embeddings | 65. natural language processing |
| 4. random forest | 35. k-means | 66. text mining |
| 5. random forests | 36. cosine similarity | 67. visual data mining |
| 6. random decision forest | 37. dbscan | 68. deep learning |
| 7. random decision forests | 38. hierarchical clustering | 69. active learning |
| 8. decision tree | 39. partitional clustering | 70. supervised learning |
| 9. naïve bayes | 40. tokenization | 71. unsupervised learning |
| 10. naive bayes | 41. topic modeling | 72. semi-supervised learning |
| 11. bayes multinomial | 42. stemming | 73. reinforcement learning |
| 12. bayesian belief network | 43. stemmer | 74. feature selection |
| 13. neural network | 44. convolution | 75. attribute selection |
| 14. convolutional neural network | 45. autoencoder | 76. variable subset selection |
| 15. CNN | 46. principal component analysis | 77. dimensionality reduction |
| 16. recurrent neural network | 47. PCA | 78. automatic term recognition |
| 17. multi-layer perceptron | 48. softmax | 79. named entity recognition |
| 18. multi-perceptron | 49. doc2vec | 80. association extraction |
| 19. semantic model | 50. word2vec | 81. automatic classification |
| 20. semantic models | 51. evo-SVM | 82. f measure |
| 21. k-nearest neighbour | 52. evoSVM | 83. f1-measure |
| 22. latent dirichlet allocation | 53. abstrackr | 84. f-score |
| 23. LDA | 54. gapscreener | 85. training set |
| 24. statistical relational learning | 55. pimienta | 86. test set |
| 25. genetic algorithm | 56. rapidminer | 87. validation set |
| 26. instance-based learning | 57. tensorflow | 88. corpus |
| 27. conditional random field | 58. scikit-learn | 89. rmse |
| 28. conditional random fields | 59. sklearn | 90. stop words |
| 29. hidden markov model | 60. weka | 91. parts of speech |
| 30. hidden markov models | 61. keras | 92. n-grams |
| 31. linear chain | 62. artificial intelligence | |

Apéndice D

Evaluación metodológica AMSTAR-2

Se enumeran a continuación únicamente los ítems de la escala metodológica AMSTAR-2 evaluados en nuestro estudio. Se excluyen así los elementos 11, 12 y 16 de la escala original.

1. ¿Incluían las preguntas de investigación y los criterios de inclusión los componentes PICO?
2. ¿Contenía el informe de la revisión una declaración explícita de que los métodos de revisión se establecieron antes de la realización de la revisión, y justificaba el informe cualquier desviación significativa del protocolo?
3. ¿Se explicó la selección de los diseños de estudio elegibles para su inclusión en la revisión?
4. ¿Los autores de la revisión utilizaron una estrategia de búsqueda bibliográfica exhaustiva?
5. ¿Los autores de la revisión realizaron la selección de estudios por duplicado?
6. ¿Los autores de la revisión realizaron la extracción de datos por duplicado?
7. ¿Se proporcionó una lista de los estudios excluidos y se justificaron las exclusiones?
8. ¿Los autores de la revisión describieron los estudios incluidos con suficiente detalle?
9. ¿Los autores de la revisión utilizaron una técnica satisfactoria para evaluar el riesgo de sesgo en los estudios individuales que se incluyeron en la revisión?
10. ¿Se informó sobre las fuentes de financiación de los estudios incluidos en la revisión?
13. ¿Los autores de la revisión tuvieron en cuenta el riesgo de sesgo en los estudios individuales al interpretar o discutir los resultados de la revisión?
14. ¿Los autores de la revisión proporcionaron una explicación satisfactoria y una discusión de cualquier heterogeneidad observada en los resultados de la revisión?
16. ¿Informaron los autores de la revisión sobre cualquier fuente potencial de conflicto de intereses, incluida cualquier financiación que hayan recibido para realizar la revisión?

Apéndice E

Normas de publicación de la revista

INTRODUCTION

The Journal of Clinical Epidemiology aims at promoting the quality of clinical epidemiologic and patient-oriented health services research through the following: Advancement and application of innovative methods of conducting and presenting primary research; Synthesizing research results; Disseminating results; And translating results into optimal clinical practice with special attention to the training of new generations of scientists and clinical practice leaders.

Contributions are welcomed from any pertinent source but should be written clearly enough to be understood by scholarly clinical readers and clinical researchers.

Authorship

All authors should have made substantial contributions to all of the following: (1) the conception and design of the study, or acquisition of data, or analysis and interpretation of data, (2) drafting the article or revising it critically for important intellectual content, (3) final approval of the version to be submitted. A description of the contribution of each individual listed as an author will be requested by the journal at the time of submission

Article Types and Lengths

The Journal of Clinical Epidemiology accepts submissions for the following article types. Please do not exceed the maximum word count or table/figure limits listed. * Reviews: 3,000 words, 50 references, 5 figures, 5 tables * Original Articles: 3,000 words, 50 references, 5 figures, 5 tables * Letters: 500 words * Commentaries: 2,500 words, 10 references, 2 figures, 2 tables. * Additional figures and tables, as well as appendices, may be included as supplementary, online-only material.

ARTICLE STRUCTURE

Formatting requirements

There are no strict formatting requirements but all manuscripts must contain the essential elements needed to convey your manuscript, for example Abstract, Keywords, Introduction, Materials and Methods, Results, Conclusions, Artwork and Tables with Captions. If your article includes any Videos and/or other Supplementary material, this should be included in your initial submission for peer review purposes.

Subdivision - numbered sections

Divide your article into clearly defined and numbered sections. Subsections should

be numbered 1.1 (then 1.1.1, 1.1.2, ...), 1.2, etc. (the abstract is not included in section numbering). Use this numbering also for internal cross-referencing: do not just refer to 'the text'. Any subsection may be given a brief heading. Each heading should appear on its own separate line.

Appendices

If there is more than one appendix, they should be identified as A, B, etc. Formulae and equations in appendices should be given separate numbering: Eq. (A.1), Eq. (A.2), etc.; in a subsequent appendix, Eq. (B.1) and so on. Similarly for tables and figures: Table A.1; Fig. A.1, etc.

Essential Title Page Information

- **Title.** Concise and informative. Titles are often used in information-retrieval systems. Avoid abbreviations and formulae where possible. JCE has adopted the editorial policy of "more informative titles" (MITs) that crisply and concisely tell our readers what our authors found in their research. A MIT states the study type and summarizes its key findings, using the past tense for individual studies and the present tense for systematic reviews. Although we are ready to assist authors in generating MITs for their work, we encourage them to submit draft MITs with their original submissions.
- **Author names and affiliations.** Where the family name may be ambiguous (e.g., a double name), please indicate this clearly. Present the authors' affiliation addresses (where the actual work was done) below the names. Indicate all affiliations with a lowercase superscript letter immediately after the author's name and in front of the appropriate address.
- **Corresponding author.** Clearly indicate who will handle correspondence at all stages of refereeing and publication, also post-publication. Ensure that phone numbers (with country and area code) are provided in addition to the e-mail address and

the complete postal address. Contact details must be kept up to date by the corresponding author.

Highlights

Highlights should be submitted in a separate editable file in the online submission system. Please use 'Highlights' in the file name and include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point).

Abstract

Each original article must have an abstract/summary not exceeding 200 words. Abstracts must be structured with the following headings: Objective, Study Design and Setting, Results, and Conclusion. An abstract is often presented separately from the article, so it must be able to stand alone. For this reason, References should be avoided, but if essential, then cite the author(s) and year(s). Also, non-standard or uncommon abbreviations should be avoided, but if essential they must be defined at their first mention in the abstract itself. Double-space abstracts, and print them on a separate page. Abstracts not in compliance with this format will be returned to the authors for revision. The bottom of the abstract page should list six key words (index-appropriate terms), a running title and a word count.

Graphical abstract

Although a graphical abstract is optional, its use is encouraged as it draws more attention to the online article. The graphical abstract should summarize the contents of the article in a concise, pictorial form designed to capture the attention of a wide readership.

Keywords

Immediately after the abstract, provide a maximum of 6 keywords, using American

spelling and avoiding general and plural terms and multiple concepts.

What is new?

- Authors are asked to provide a text box entitled "What is new" that summarized what this paper adds to the existing literature, with up to 5 items/points addressing the following areas: Key findings; What this adds to what is known; What is the implication, what should change now.

REFERENCES

Citation in text

Please ensure that every reference cited in the text is also present in the reference list (and vice versa). Any references cited in the abstract must be given in full. Unpublished results and personal communications are not recommended in the reference list, but may be mentioned in the text. If these references are included in the reference list they should follow the standard reference style of the journal and should include a substitution of the publication date with either 'Unpublished results' or 'Personal communication'. Citation of a reference as 'in press' implies that the item has been accepted for publication.

Reference links

Increased discoverability of research and high quality peer review are ensured by online links to the sources cited. In order to allow us to create links to abstracting and indexing services, such as Scopus, CrossRef and PubMed, please ensure that data provided in the references are correct. Please note that incorrect surnames, journal/book titles, publication year and pagination may prevent link creation. When copying references,

please be careful as they may already contain errors. Use of the DOI is highly encouraged.

Web references

As a minimum, the full URL should be given and the date when the reference was last accessed. Any further information, if known (DOI, author names, dates, reference to a source publication, etc.), should also be given. Web references can be listed separately (e.g., after the reference list) under a different heading if desired, or can be included in the reference list.

Data references

This journal encourages you to cite underlying or relevant datasets in your manuscript by citing them in your text and including a data reference in your Reference List. Data references should include the following elements: author name(s), dataset title, data repository, version (where available), year, and global persistent identifier. Add [dataset] immediately before the reference so we can properly identify it as a data reference. The [dataset] identifier will not appear in your published article.

Reference formatting

There are no strict requirements on reference formatting at submission. References can be in any style or format as long as the style is consistent. Where applicable, author(s) name(s), journal title/book title, chapter title/article title, year of publication, volume number/book chapter and the article number or pagination must be present. Use of DOI is highly encouraged. The reference style used by the journal will be applied to the accepted article by Elsevier at the proof stage.