VALIDEZ Y CONCORDANCIA DEL APRENDIZAJE DE MÁQUINAS EN LA EVALUACIÓN DE RIESGO DE SESGOS DE ENSAYOS CLÍNICOS ALEATORIZADOS. REVISIÓN SISTEMÁTICA

Autor:
Heider Alexis Bautista Mier

Maestría en Métodos para la producción y aplicación del conocimiento científico en salud
Universidad Autónoma de Bucaramanga
Facultad de ciencias de la salud
2022

VALIDEZ Y CONCORDANCIA DEL APRENDIZAJE DE MÁQUINAS EN LA VALUACIÓN DE RIESGO DE SESGOS DE ENSAYOS CLÍNICOS ALEATORIZADOS. REVISIÓN SISTEMÁTICA

Autor:
Heider Alexis Bautista Mier

Trabajo de grado para optar por el título de Magíster en métodos para la Producción y aplicación de conocimiento científico en salud.

Directora:
Skarlet Marcell Vásquez Hernández,
Enf, Magister en Epidemiología Clínica

Asesores temáticos:
Alexis Barreto Montenegro
Bibliotecólogo

Karen Julieth Moreno
Ft, Magister en Epidemiología

Maestría en Métodos para la producción y aplicación del conocimiento científico en salud
Universidad Autónoma de Bucaramanga
Facultad de ciencias de la salud
2022

Tabla de contenido

Lista de tablas

Lista de Figuras

1. TÍTULO DEL PROYECTO

Validez y concordancia del aprendizaje de máquinas en la automatización de la evaluación de riesgo de sesgos de ensayos clínicos aleatorizados durante el desarrollo de revisiones sistemáticas. Revisión sistemática

2. RESUMEN DEL PROYECTO

Introducción: La evaluación del riesgo de sesgo de ensayos clínicos es una actividad crítica en el desarrollo de revisiones sistemáticas. El aprendizaje de máquinas podría disminuir la variabilidad y subjetividad inherente a este proceso.

Objetivo: Determinar la validez y concordancia del aprendizaje de máquinas en la automatización de la evaluación de riesgos de sesgos de estudios clínicos primarios incluidos en revisión sistemática.

Metodología: Se realizó una revisión sistemática que incluyó estudios que evaluaron la validez y/o concordancia de la evaluación del riesgo de sesgos de ensayos clínicos a través de herramientas de aprendizaje de máquinas en comparación con la evaluación convencional realizada por humanos. La búsqueda fue realizada en Medline, SCOPUS, LILACS, Springer Link, scienceDirect y Google scholar. La selección y evaluación del riesgo de sesgo de los estudios fue realizado por dos investigadores de manera independiente, los desacuerdos fueron resueltos mediante consenso. Los resultados se presentaron mediante una síntesis cualitativa. La síntesis cuantitativa de la información se realizó como ejercicio académico en el software Metadisc 2, se estimaron sensibilidad y especificidad como medidas resumen.

Resultados: A través de la estrategia de busqueda se identificaron 257 estudios. Una vez revisados títulos y resumen se seleccionaron 17 para lectura completa y finalmente 5 estudios de corte transversal fueron incluidos en la síntesis cualitativa. La herramienta de aprendizaje de máquinas usada con mayor frecuencia fue RobotReviewer, que usa los dominios de la herramienta de evaluación de sesgos de ROB1 (Cochrane). En general, el riesgo de sesgo de los estudios incluidos en esta revisión fue clasificado como bajo. La concordancia de la herramienta de aprendizaje de máquinas versus la evaluación por humanos fue baja; el coeficiente kappa para los dominios de generación de sencuencia aleatoria y asignación oculta osciló entre 0.40- 0.60, mientras que para los dominios de enmascaramiento del personal y participantes, y ocultamento del desenlace varió entre 0.04- 0.43. La sensibilidad y especificidad de la herramienta automatizada fue evaluada en 3 estudios; para el dominio de generación de secuencia aleatoria y asignación oculta la sensibilidad varió entre 77- 88% y la especificidad entre 62- 84%. Para los dominios de enmascaramiento del personal y los participantes, y ocultamiento de desenlaces la sensibilidad de la herramienta automatizada osciló entre 10- 47%, y la especificidad 48 y 95%. La estimación global del riesgo de sesgo fue evaluada únicamente 2 estudios, donde se reportó sensibilidad de la herramienta automatizada entre 0 a 33% y especificidad entre 95 a 98%.

Conclusión: El aprendizaje de máquinas podría ser una herramienta de apoyo para la evaluación de riesgo de sesgo de ensayos clínicos, principalmente al descartar ensayos clínicos de alto riesgo de sesgo. Por el momento, la evaluación por humanos sigue siendo indispensable. Es deseable mejorar el rendimiento de estas herramientas para implementar su uso de manera rutinaria en el desarrollo de revisiones sistemáticas.

ABSTRACT

Introduction: The evaluation of the risk of bias in clinical trials is a critical activity in the development of systematic reviews. Machine learning could reduce the variability and subjectivity inherent in this process.

Objective: To determine the validity and consistency of machine learning in the automation of bias risk assessment of primary clinical studies included in a systematic review.

Methodology: A systematic review was carried out that included studies that evaluated the validity and/or concordance of the assessment of the risk of bias of clinical trials through machine learning tools in comparison with the conventional evaluation carried out by humans. The search was carried out in Medline, SCOPUS, LILACS, Springer Link, scienceDirect and Google scholar. The selection and evaluation of the risk of bias of the studies was carried out by two researchers independently, disagreements were resolved by consensus. The results were presented through a qualitative synthesis. The quantitative synthesis of the information was carried out as an academic exercise in the Metadisc 2 software, sensitivity and specificity were estimated as summary measures.

Results: Through the search strategy, 257 studies were identified. Once the titles and abstract were reviewed, 17 were selected for full reading and finally 5 cross-sectional studies were included in the qualitative synthesis. The most frequently used machine learning tool was RobotReviewer, which uses ROB1 (Cochrane) bias assessment tool domains. In general, the risk of bias of the studies included in this review was classified as low. Concordance of machine learning tool versus human assessment was low; the kappa coefficient for the random sequence generation and allocation concealment domains ranged from 0.40-0.60, while for the staff and participant masking and outcome concealment domains it ranged from 0.04-0.43. The sensitivity and specificity of the automated tool was evaluated in 3 studies; For the domain of random sequence generation and allocation concealment, the sensitivity varied between 77-88% and the specificity between 62-84%. For the blinding of participants and personnel, and blinding of outcome assessment domains, the sensitivity of the automated tool ranged between 10% and 47%, and the specificity between 48% and 95%. The global estimation of the risk of bias was evaluated in only 2 studies, where sensitivity of the automated tool was reported between 0 to 33% and specificity between 95 to 98%.

Conclusion: Machine learning could be a support tool for evaluating the risk of bias in clinical trials, mainly by ruling out clinical trials with a high risk of bias. For the time being, human evaluation remains indispensable. It is desirable to improve the performance of these tools to implement their routine use in the development of systematic reviews.

3. DESCRIPCIÓN DEL PROYECTO

3.1 Planteamiento del problema de investigación y su justificación en términos de necesidades y pertinencia

En la toma de decisiones en salud los profesionales requieren de la gestión y consumo de grandes volúmenes de información procedente de estudios observacionales y/o ensayos clínicos controlados aleatorizados publicados en revistas científicas. Año a año el número de publicaciones en las bases de datos crece constantemente, se estima que existen más de 30 millones de publicaciones solo en PubMed (1).

Existe un acumulo de información científica que supera nuestras capacidades de análisis, por tanto, se han desarrollado metodologías para la síntesis de la mejor evidencia disponible conocidas como revisiones sistemáticas de la literatura (2,3).

Las revisiones sistemáticas (RS) se caracterizan por una búsqueda sistemática, reproducible y exhaustiva de toda la evidencia disponible, busca reunir la mayor cantidad de artículos que respondan la pregunta de investigación y posterior a evaluar la calidad de la información contenida intenta sintetizar la información en un estimado puntual, de esta manera mejora la precisión y permite incrementar la fuerza de las recomendaciones para la práctica clínica (4,5)

Estos diseños de investigación no están exentos de la introducción de errores sistemáticos durante su desarrollo, por tanto, para obtener de ellos información válida se requiere metodología explícita y compleja en la que participan un número importante de investigadores en los diferentes pasos que conlleva el desarrollo de las revisiones sistemáticas (6,7).

Consecuentemente, la producción eficiente y de alta calidad de revisiones sistemáticas exige replantear la forma de cómo se llevan a cabo. Gracias a los avances en las ciencias computacionales en los últimos 40 años se ha propuesto incorporar la automatización en los diferentes procesos de su diseño (8).

El advenimiento de la ciencia de datos, la big data y la inteligencia artificial ha permitido mejorar la velocidad del procesamiento de grandes volúmenes de información (9–11). El avance y las mejoras de los componentes de la Inteligencia artificial como machine learning, redes neuronales, Deep learning, Natural language processing (NLP), minería de textos han permitido superar a los softwares y sistemas computacionales tradicionales, facilitando la síntesis y transferencia de los datos, generando gran potencial para aplicarse en la producción de investigación secundaria (12–17).

El procesamiento de lenguaje natural (NLP) se presenta como una herramienta útil para la búsqueda, clasificación y selección de artículos presentes en las bases de datos de investigación biomédica, ya que permite la evaluación de títulos y resúmenes gracias a su capacidad de realizar vinculación gramática, evaluar similitud semántica, aceptabilidad lingüística y realiza razonamiento de sentido común. De esta manera esta tecnología logra reducir de manera importante la carga de trabajo de los investigadores a la hora de realizar

interés en la búsqueda activa y selección de estudios, sin pérdidas considerables de artículos de investigación (18–20)

Por su parte, la minería de texto utiliza varias técnicas y herramientas que se utilizan para detectar patrones de texto y permitir extraer información partir de métodos estadísticos, al implementarse como herramienta de búsqueda apoyar la extracción de información en revisiones sistemáticas, principalmente aquellas que no requieran la máxima sensibilidad como revisiones rápidas (21–24). Entre los componentes de la inteligencia artificial el aprendizaje de máquinas se caracteriza por su versatilidad, pudiendo participar e integrarse en los diferentes pasos del desarrollo de revisiones sistemáticas (25).

La búsqueda de la literatura es uno de los pasos cruciales en el desarrollo de las RS, en este aspecto el aprendizaje de máquinas permite realizar screening de manera rápida de títulos y resúmenes a través de análisis automático de texto libre, reduciendo el tiempo y facilitando el trabajo de los investigadores (20,26).

La selección de los estudios durante la conducción de las RS a través de algoritmos de aprendizaje de máquinas permite mejorar la eficiencia con la cual se desarrollan. Ya que disminuye el número necesario de revisores y el tiempo necesario para realizar este proceso, sin impactar de manera significativa en su calidad y a menor costo (27,28). El proceso de extracción de la información mediante la técnica de aprendizaje de máquinas permite la extracción rápida de información (29,30).

De otro lado, la evaluación de sesgos de los estudios primarios que agrupan las revisiones sistemáticas es un componente crítico ya que permite establecer la idoneidad y veracidad de las estimaciones individuales de cada estudio. Los dominios que pueden influir en la idoneidad de los ensayos clínicos son la generación de la secuencia aleatoria, la asignación oculta, el enmascaramiento de participantes e investigadores, el enmascaramiento del desenlace, la presencia de datos incompletos, el reporte selectivo, por tanto, se han diseñado modelos automatizados para evaluar el riesgo de sesgo presente en los estudios a través de la valoración parcial o total estos dominios (31,32). Sin embargo, la precisión de las evaluaciones realizadas por las herramientas automatizadas aún es incierta y por lo tanto, su utilidad es limitada.

3.1.1 Planteamiento del problema

Las revisiones sistemáticas tienen un papel importante al facilitar a profesionales de la salud la toma de decisiones, mediante estas se puede realizar una búsqueda sistemática, reproducible y exhaustiva de toda la evidencia disponible, reúne la mayor cantidad de información procedente de investigación primaria que responda la pregunta de investigación, evalúa la calidad de la información contenida e intenta resumir la información en un estimado puntual, mejorando la precisión y permitiendo incrementar la fuerza de las recomendaciones para la práctica clínica (3–5).

Las revisiones sistemáticas tienen como objeto de estudio tanto estudios observacionales como experimentales, estos últimos de gran interés ya que se consideran el diseño más robusto para la evaluación de los efectos de las intervenciones clínicas (4). No obstante, estos tipos de estudios no están exentos de la introducción de errores sistemáticos durante su desarrollo, llegando a influir en los resultados de la investigación al sobreestimar o

subestimar el verdadero efecto de las intervenciones. Por tanto, la síntesis de información de calidad requiere de la evaluación del riesgo de sesgos de los estudios incluidos para establecer la confianza en torno al efecto estimado sobre la población.

Aunque puede ser una tarea compleja para los investigadores establecer la presencia de sesgos y su impacto en un estudio, se han desarrollado herramientas que permiten a los investigadores reconocer ciertas fallas durante el diseño, como la aleatorización, la asignación de la intervención oculta y el enmascaramiento del personal, participantes o desenlaces que podrían afectar los resultados. La colaboración Cochrane ha desarrollado su instrumento de referencia para la evaluación del riesgo de sesgo para ser realizada por dos investigadores de manera independiente y en caso de desacuerdo recomienda resolver por medio de consenso o un tercer investigador. Inicialmente la herramienta propuesta fue ROB1 para la evaluación del riesgo de sesgo (33), las encuestas realizadas a los usuarios documentaron que el uso de esta herramienta mejoró la calidad de la evaluación del riesgo de sesgo, alcanzado una la aceptabilidad de la herramienta superior al 80% (34).

Sin embargo a pesar de la estandarización, con el tiempo se documentó una elevada prevalencia de un inadecuado juicio de dicho riesgo de sesgos por investigadores (35–37), adicionalmente algunos estudios han sugerido que esta herramienta presenta una confiabilidad moderada para establecer un juicio sobre riesgo de sesgo (38).

Como respuesta a estas dificultades, se desarrolló una actualización de la versión inicial (ROB 2), la cual establece un marco para evaluar el riesgo de sesgo, siendo los dominios más importantes mediante los cuales se puede introducir sesgo en los resultados de un ensayo, en función de un combinación de evidencia empírica y consideraciones teóricas, incluyendo la evaluación de sesgo derivado del proceso de aleatorización, sesgo debido a desviaciones de las intervenciones previstas, sesgo debido a la falta de datos de resultados, sesgo en la medición del resultado, sesgo en la selección del resultado informado (39,40).

El aprendizaje de máquinas podría reducir el esfuerzo humano y variabilidad inherente en esta tarea, los datos iniciales sugieren un adecuado rendimiento que se ha observado de la incorporación de esta tecnología en el proceso de la evaluación de riesgo de sesgos (31,32).

3.2 Pregunta de investigación

¿Cuál es la validez y concordancia de las herramientas automatizadas de aprendizaje de máquinas comparado con la estrategia convencional de revisores humanos para la evaluación del riesgo de sesgo de ensayos clínicos durante el desarrollo de una revisión sistemática?

Propósito: Diagnóstico de riesgo de sesgo

4. MARCO TEÓRICO

Las revisiones sistemáticas son importantes para la toma de decisiones tanto en políticas de salud como en la práctica asistencial, por lo tanto, lograr su desarrollo en menor tiempo y consumo de recursos económicos y humanos es crucial. La integración de la automatización en los procesos de desarrollo de revisiones sistemáticas podría facilitar la síntesis de la información científica con mayor eficiencia.

Para comprender el alcance del objetivo que pretende alcanzar esta propuesta de investigación consideramos introducir las definiciones, utilidad, fortaleza y limitaciones de cada uno de los términos centrales de este trabajo de investigación: Ensayo clínico aleatorizado, sesgos, revisión sistemática, aprendizaje de máquinas.

Ensayo clínico aleatorizado (ECA)

Los ECA son un tipo de diseño de investigación clínico prospectivo, siguen un protocolo pre-especificado cuyo objetivo es evaluar una intervención entre un grupo experimental y uno o varios controles, utilizan estrategias metodológicas que permiten que estos grupos sean comparables permitiendo establecer si los resultados se asocian o no al factor experimental. Los ECA representan la mejor evidencia para evaluar la eficacia, tolerabilidad, seguridad de una intervención para una condición clínica particular (41,42).

En medicina basada en la evidencia los ECA se consideran jerárquicamente superiores a otros tipos de estudios como los de cohortes, casos y controles, series de casos debido a su validez y confiabilidad (43). Para establecer que un ECA es alta calidad se requiere un bajo riesgo de sesgos.

Los sesgos

Los sesgos son conocidos como errores o desviaciones de la verdad de manera sistemática, pueden sobrevalorar o infravalorar la magnitud y dirección del efecto de una intervención y conducen a resultados inválidos (44). Estos se pueden introducir durante el diseño, la inscripción de participantes, recopilación de datos, el análisis y la elaboración de los informes (45).

Pueden clasificarse a grandes rasgos en sesgos de selección, de confusión y de información. El sesgo de selección hace referencia a error sistemático que se presenta al incluir participantes en el estudio, también se presenta cuando los estudios publicados constituyen una muestra imparcial de los estudios realizados. El sesgo de información se produce durante la recopilación de datos. El error se presenta cuando el proceso para detectar el estado de exposición y/o el resultado se encuentra alterado (46,47).

El sesgo de confusión es una desviación de la estimación en un estudio producida por la distribución desigual o desequilibrada en los grupos de comparación de una tercera variable. Si esta variable de confusión es un predictor del efecto, entonces su distribución desigual contaminará la verdadera relación entre la exposición y el efecto o resultado evaluado (48).

La fuente de sesgos en estudios clínicos se ha evaluado mediante análisis de estudios meta-epidemiológicos. Savovic y colaboradores, documentaron que el efecto de una intervención puede sobreestimarse cuando se realiza de manera inadecuada la aleatorización de los participantes o en el enmascaramiento de los participantes y los investigadores (49).

Teniendo en cuenta el impacto negativo que pueden tener los sesgos al momento de tomar decisiones sobre intervenciones en salud, desde el año 2005 la colaboración Cochrane ha trabajado en el desarrollo de herramientas que permitan establecer el riesgo de sesgo presente en las diferentes etapas del desarrollo de los estudios clínicos de intervención. En su primera versión, introducida en 2008 (ROB 1) incluye la evaluación del riesgo de selección al identificar la generación de secuencia aleatoria y asignación oculta, sesgo de rendimiento al establecer el enmascaramiento de los participantes e investigadores, sesgo de detección el determinar el cegamiento de la evaluación de resultados, sesgo de deserción al identificar la presencia de datos de los desenlaces incompletos, sesgo de reporte al establecer la presencia de informe selectivo (33).

Después de casi una década de uso de la herramienta RoB y tras documentarse falencias que podrían poner en riesgo la evaluación del riesgo de sesgo de ensayos clínicos la colaboración Cochrane realizó una actualización de la versión inicial (ROB 2), la cual establece un marco para evaluar el riesgo de sesgo, siendo los dominios más importantes mediante los cuales se puede introducir sesgo en los resultados de un ensayo, en función de un combinación de evidencia empírica y consideraciones teóricas, incluyendo la evaluación de sesgo derivado del proceso de aleatorización, sesgo debido a desviaciones de las intervenciones previstas, sesgo debido a la falta de datos de resultados, sesgo en la medición del resultado, sesgo en la selección del resultado informado (39,40).

Revisión sistemática y meta-análisis

Las revisiones sistemáticas mejoran la toma de decisiones en la práctica clínica ya que utilizan la información procedente de la mejor evidencia disponible, gracias a su rigurosa, transparente y reproducible metodología (50).

Las revisiones sistemáticas son diseños de estudios que permiten sintetizar la información obtenida de estudios primarios en torno a una pregunta particular, utilizando una metodología sistemática, lo cual le garantiza transparencia y reproducibilidad. Para su realización requiere la pregunta de investigación, establecer la estructura de búsqueda de la literatura, determinar los criterios de selección de las fuentes de información, seleccionar los artículos, establecer la calidad de los artículos seleccionados y el riesgo de sesgo, extraer la información allí depositada, analizar los datos obtenidos y realizar la síntesis (51,52).

Para completar una revisión sistemática los investigadores pueden tardar hasta 2 años, siendo un proceso complejo y demandante, por tanto, para su desarrollo se han establecido alternativas como desarrollo de herramientas de colaboración colectiva en línea y la automatización (53).

Aprendizaje de máquinas

El aprendizaje de máquinas (AM) es una área de las ciencias computacionales que busca el "aprendizaje de tareas" de las computadoras para su posterior desarrollo de manera autónoma, utiliza grandes volúmenes de datos como su fuente de información, esta información y mejora su desempeño en la medida que el modelo adquiere "experiencia". Los algoritmos de aprendizaje automático utilizan 3 métodos de "aprendizaje": supervisado, no supervisado y semi-supervisado. En el aprendizaje supervisado, el resultado (es decir, la variable dependiente o "etiqueta") se conoce, alimentado con datos estructurados para los atributos (es decir, las variables independientes o "características"), el algoritmo intenta encontrar el modelo correspondiente que predice los resultados del paciente con la mayor precisión, exactitud o recuperación. El AM supervisado utiliza la regresión lineal y logística. Para el aprendizaje no supervisado el algoritmo intenta establecer relaciones entre las características de los pacientes sin conocer los resultados para agrupar a los datos en función de sus similitudes. Finalmente, el semi-supervisado ajusta los modelos a los datos etiquetados y no etiquetados. Etiquetar los datos (resultados) a menudo lleva mucho tiempo y es costoso, especialmente para grandes conjuntos de datos. El aprendizaje semi-supervisado complementa los datos etiquetados limitados con una gran cantidad de datos no etiquetados con el objetivo de mejorar el rendimiento del modelo (54). Los algoritmos utilizados en AM más frecuente son las redes neuronales artificiales, árbol de decisión, máquinas de vectores de soporte, naive Bayes y k-means clustering.

Las redes neuronales tienen la capacidad para acomodar interacciones variables y asociaciones no lineales sin especificación del usuario. Los árboles de decisión crean una serie de reglas de decisión basadas en variables de entrada continuas y/o categóricas para predecir un resultado, Support vector machines (SVMs) son un conjunto de métodos de aprendizaje supervisado que se utilizan para problemas de clasificación y regresión. Las SVM generalmente demuestran un bajo error de clasificación errónea y escalan bien a datos de alta dimensión, un algoritmo de naive Bayes es un algoritmo de clasificación probabilística simple basado en el teorema de Bayes que hace la suposición "naive" de independencia entre variables predictivas, calcula la probabilidad asociada con cada clase posible condicionada a un conjunto de covariables, es decir, el producto de la probabilidad previa y la función de probabilidad. El clasificador luego selecciona la clase con la probabilidad más alta como la clase "correcta" (55,56).

5. ESTADO DEL ARTE

La medicina basada en la evidencia se basa en utilizar la mejor información disponible, involucrando los valores de los pacientes y la experiencia clínica para la toma de decisiones clínicas. Las revisiones sistemáticas son la piedra angular en la toma de decisiones ya que se caracterizan por ser un método que permite reunir y sintetizar la información disponible en torno a una pregunta específica. No obstante, su desarrollo y actualización constante se ha convertido en un verdadero reto, dada la alta carga de trabajo especializado que demanda, por tanto, se han introducido métodos de automatización para su desarrollo (57).

Actualmente, algunas tareas del desarrollo de revisiones sistemáticas se llevan a cabo de manera automatizada (58). Cowie et al., publicó en 2022 una revisión sistemática en torno a este tópico e identificó 24 softwares que facilitan el desarrollo de la investigación secundaria, el 63% de las herramientas identificadas facilitan la búsqueda, extracción y exportación de la información, 21% apoyan solo la etapa de cribado (59).

Por su parte, Scott et al. realizó una encuesta a 253 investigadores con experiencia en producción de revisiones sistemáticas identificando que el 89% ha utilizado herramientas de automatización, de estos el 79% las utiliza para el "screening", siendo herramientas más utilizadas Covidence, RevMan, Rayyan y GRADEPro. Adicionalmente, el 80% de los encuestados describen que estas herramientas permiten ahorrar tiempo y el 54% consideran que mejoran la precisión (60).

Incorporar estas herramientas podrían facilitar a los grupos investigadores el desarrollo de información potencialmente útil para los tomadores de decisiones, datos aportados por Clark et al., sugieren que utilizando la automatización se podría completar el desarrollo de una revisión sistemática en un periodo de 2 semanas (61).

Una de las tareas críticas de las revisiones sistemáticas y que incrementa la confianza de la información obtenida a través de ella es la evaluación del riesgo de sesgo. Para realizar dicha tarea se recomienda contar con 2 investigadores que participen de manera independiente en su realización. En los últimos años se ha incluido la automatización para facilitar su desarrollo (31,32). Dada la importancia que tiene esta tarea se ha considerado vital establecer validez y/o concordancia del uso de aprendizaje de máquinas comparado con la estrategia convencional en la evaluación de riesgo de sesgo.

A la fecha, no existen revisiones sistemáticas que resuman la evidencia disponible en torno a la automatización de la evaluación del riesgo de sesgos en ensayos clínicos controlados.

6. OBJETIVOS

6.1 Objetivo general

Determinar la validez y concordancia del aprendizaje de máquinas en la automatización de la evaluación de riesgos de sesgos de ensayos clínicos en comparación a la evaluación realizada por revisores humanos durante el desarrollo de una revisión sistemática de la literatura.

6.2 Objetivos específicos

Identificar los diferentes softwares que utilizan el aprendizaje de máquinas en la evaluación de riesgo de sesgo comparado con la evaluación por revisores humanos durante el desarrollo de una revisión sistemática de la literatura.

Describir las características de los estudios que evaluaron la validez y concordancia de herramientas de aprendizaje de máquinas para la evaluación de riesgo de sesgos de ensayos clínicos aleatorizados comparado con la evaluación por revisores humanos durante el desarrollo de revisiones sistemáticas.

Determinar la concordancia del aprendizaje de máquinas para la evaluación de riesgo de sesgos de ensayos clínicos comparado con la evaluación por revisores humanos durante el desarrollo de una revisión sistemática de la literatura.

Determinar la validez de los diferentes softwares de aprendizaje de máquinas en la evaluación de riesgo de sesgo de ensayos clínicos comparado con la evaluación por revisores humanos durante el desarrollo de revisiones sistemáticas.

7. METODOLOGÍA

Diseño del estudio: Se realizó una revisión sistemática de la literatura. El protocolo se publicó previo a su desarrollo en el siguiente enlace: https://repository.unab.edu.co/handle/20.500.12749/15214

7.4.1 Criterios de elegibilidad

Se incluyeron en esta revisión sistemática todos los estudios que compararon el uso de aprendizaje de máquinas con la estrategia convencional en la evaluación de riesgo de sesgo de ensayos clínicos (revisión por humanos) y que establecieron la validez y/o concordancia y/o precisión del aprendizaje de máquinas.

En ese sentido, se consideró la siguiente estructura para establecer la busqueda.

Población	Estudios diagnósticos que incluyan evaluación de riesgo de sesgos en ensayos clínicos
Tipos de intervención	Herramientas automatizadas de aprendizaje de máquinas
Tipos de comparador	Estrategia convencional de evaluación por humanos
Tipos de medidas de desenlace	Validez y concordancia de la evaluación de riesgo de sesgos

Los desenlaces de interés fueron definidos como:

- Validez: Se determinó en base a los valores de sensibilidad, especificidad y exactitud con respecto a la evaluación realizada por revisores humanos.
- Concordancia: Se definió como acuerdo de la clasificación del riesgo de sesgo realizada por la herramienta de aprendizaje de máquinas y los revisores humanos. Determinado por el índice Kappa, valores de 0 se considera sin acuerdo, 0,00-0,20 el grado de acuerdo es insignificante, 0,21-0,40 el grado de acuerdo es mediano, 0,41-0,60 el grado de acuerdo es moderado, 0,61-0,80 el grado de acuerdo es sustancial, 0,81-1,00 el grado de acuerdo es casi perfecto.

Se determinó incluir inicialmente revisiones sistemáticas, en ausencia de estas, estudios observacionales. Se incluyeron estudios primarios en inglés y español, publicados hasta el 28 de febrero de 2022.

7.4.2 Criterios de exclusión

Se excluyó revisiones narrativas, editoriales o estudios observacionales que evaluaron el riesgo de sesgo mediante otro componente diferente al aprendizaje de máquinas.

7.1 Búsqueda de la información

La búsqueda fue realizada por 3 investigadores de forma independiente, siguiendo la ecuación establecida para cada una de las bases de datos.

Se realizó la búsqueda en Scopus, ScienceDirect, Medline, Lilacs, Springer link y para identificar artículos en la literatura gris se usó Google Scholar. Debido al acceso restringido, la búsqueda no pudo ser replicada en EMBASE.

La Tabla 1. Presenta los términos de búsqueda que se usaron para cada una de las bases de datos:

Tabla 1. Términos de búsqueda y estrategia de búsqueda

Base de datos	**Ecuación de búsqueda**
Scopus (Acceso por suscripción)	INDEXTERMS (machine AND learning) OR KEY (machine AND learning) AND KEY (bias) OR INDEXTERMS (risk of bias) AND KEY (automation)
Medline (Acceso abierto)	("machine learning"[MeSH Terms] OR "machine learning system"[All Fields]) AND ("bias"[MeSH Terms] OR "bias"[All Fields]) AND ("clinical trial"[Publication Type] OR "clinical trials as topic"[MeSH Terms] OR "clinical trials"[All Fields])
Lilacs (Acceso abierto)	systematic review automation bias evaluation
Springer Link	(“Machine learning”) AND (“Clinical Trial”) AND ("systematic review") AND ("Risk of bias assessment") AND (Validation OR Studies OR Usefulness OR Utility OR Reliability OR Accuracy)
ScienceDirect	"Machine learning" AND "risk of bias" AND "systematic review" AND "automation"
Google Scholar (Acceso abierto)	(“machine learning system”) AND (“Clinical Trial”) AND ("systematic review") AND ("Risk of bias assessment") AND (Validation OR Studies OR Usefulness OR Utility OR Reliability OR Accuracy)

7.4 Selección de los estudios

La selección fue realizada por dos investigadores de manera independiente. Esta fue realizada en dos pasos; en el primero los investigadores realizaron la lectura de los títulos y resúmenes, en el segundo se realizó una lectura en profundidad del texto completo de los estudios seleccionados en el primer paso, con el fin de identificar los estudios potencialmente elegibles.

Los desacuerdos en la selección de los estudios en el primer o segundo paso entre los dos investigadores fueron resueltos mediante discusión y consenso.

Los resultados del proceso se resumieron siguiendo el flujograma PRISMA. El proceso de selección de la información fue documentado mediante tablas para la gestión de los datos obtenidos a lo largo de la revisión.

7.5 Extracción de la información

La extracción de la información fue realizada por un investigador, quien sustrajo los datos relacionados con la metodología de los estudios incluidos, las características de la intervención y del comparador, y los desenlaces identificados.

Adicionalmente se documentaron las características bibliométricas de los estudios incluidos: idioma, revista, factor de impacto, país de publicación y el año de publicación.

7.6 Evaluación del riesgo de sesgos de los estudios incluidos en la revisión

La evaluación del riesgo de sesgo de los estudios incluidos fue realizada por dos investigadores de manera independiente y las discrepancias se resolvieron por consenso. El riesgo de sesgo de los estudios incluidos (orientados a pruebas diagnósticas) se evaluó usando el instrumento Critical Appraisal Skills Programme, versión en español (CASPe). Los hallazgos de la evaluación fueron validados aplicando posteriormente el instrumento QADAS2.

El riesgo de sesgo fue clasificado como “riesgo bajo” o “riesgo alto” o “riesgo incierto” y se presentan los resultados de esta evaluación en una tabla de resumen de acuerdo a la tipología de cada estudio.

7.7 Síntesis cualitativa y cuantitativa.

La presentación del documento se realizó siguiendo las recomendaciones PRISMA. Para el análisis de la información se realizó una síntesis cualitativa de los estudios, que se presentó en tablas resumen.

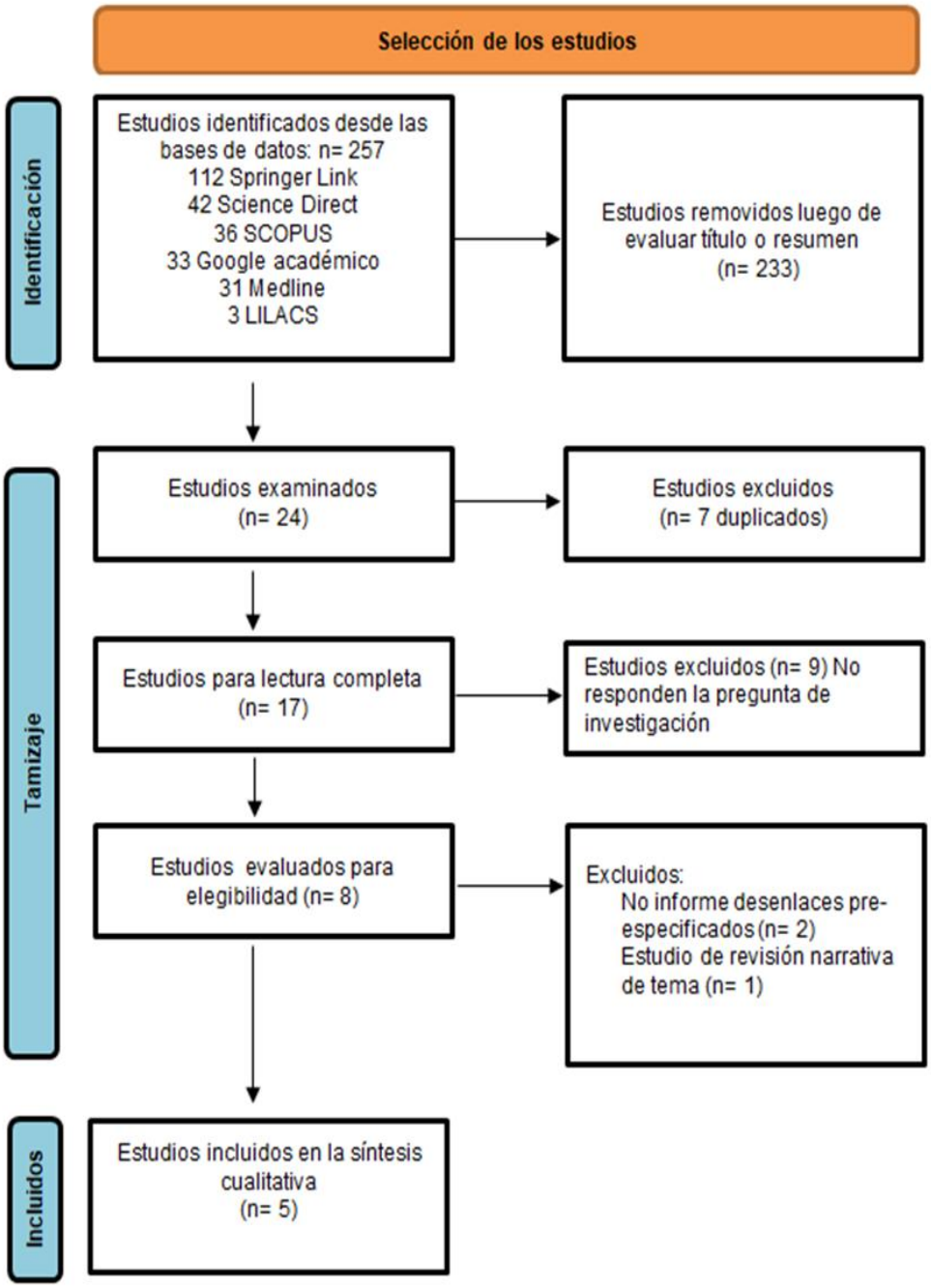
Se estableció a priori realizar síntesis cuantitativa de la información en ausencia de efecto umbral (I2<40%), y utilizando el método de efectos aleatorios con el modelo bivariado. Sin embargo, como parte del ejercicio académico del investigador principal, pese a que los hallazgos no cumplían con los criterios para realizar metaanálisis, este fue ejecutado.

Para el análisis se utilizó el programa software Metadisc2, el cual es un software estadístico de acceso libre online para la realización de metaanálisis de pruebas diagnósticas (https://ciberisciii.shinyapps.io/MetaDiSc2/).

8. RESULTADOS

La aplicación de los términos de búsqueda en las bases de datos identificó 257 estudios. Posterior a la revisión del título y resumen, 24 artículos fueron seleccionados para revisión del texto completo. Después de aplicar los criterios de exclusión, 5 estudios fueron seleccionados para la síntesis cualitativa y 3 incluidos para la síntesis cuantitativa (31,62–65) (Ver Figura 1).

Figura 1. Flujograma proceso de búsqueda y selección de artículos



La Tabla 2 presenta las características bibliométricas de los estudios incluidos. Los estudios se publicaron entre el 2015 y el 2021, principalmente en Canadá y Reino Unido.

Tabla 2. Resumen de las características bibliométricas de los estudios a incluir.

Ref,	Título	Idioma	Revista	Factor de impacto	País de publicación	Año de publicación
(62)	Agreement in Risk of Bias Assessment Between RobotReviewer and Human Reviewers: An Evaluation Study on Randomized Controlled Trials in Nursing-Related Cochrane Reviews	Inglés	Journal of Nursing Scholarship	3,176	Switzerland.	2021
(63)	Technology-assisted risk of bias assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool.	Inglés	Journal of Clinical Epidemiology	6,437	Canadá	2018
(64)	Automating Risk of Bias Assessment for Clinical Trials.	Inglés	ieee journal of biomedical and health informatics	5,772	Reino Unido	2015
(65)	Comparing machine and human reviewers to evaluate the risk of bias in randomized controlled trials.	Inglés	research synthesis and methods	5,273	Canadá	2020
(31)	RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials.	Inglés	journal of the american medical informatics association	4,112	Reino Unido	2016

La Tabla 3 presenta las características de los estudios incluidos que midieron como desenlaces validez y concordancia (62, 63, 65), mientras que la Tabla 4 presenta las características de los estudios que midieron como desenlace la validez global (64, 31). Todos correspondían a diseños de corte transversal, en su mayoría de temporalidad mixta. En total, estos estudios evaluaron el riesgo de sesgo mediante una herramienta de aprendizaje de máquinas en 16750 ensayos clínicos aleatorizados. La herramienta que se reportó con mayor frecuencia fue el software RobotReviewer, identificándose en 4 de los 5 estudios incluidos. RobotReviewer, es un modelo semisupervisado que extrae información identificando oraciones o palabras, procedentes de un documento específico. El modelo ha

sido programado de acuerdo con los criterios establecidos por la herramienta de evaluación de sesgos ROB 1 (Risk of Bias tool de Cochrane).

De igual manera, todos los estudios reportaron que la herramienta de evaluación de sesgos utilizada por los revisores humanos (comparador) fue ROB 1 (Risk of Bias tool de Cochrane). Sin embargo, los reportes se limitaron a 4 dominios principales: generación de la secuencia aleatoria, ocultamiento de la asignación, enmascaramiento del personal y participantes y ocultamiento del desenlace.

Los estudios reportados por Marshall et al. (64, 31), se identificaron como los primeros estudios que reportaron la validez de un modelo de aprendizaje de máquinas automatizado para evaluar el riesgo de sesgo en ensayos clínicos. El modelo reportado en el estudio en el 2015, fue la primera versión construida como una herramienta semisupervisada, con capacidad de identificar frases o palabras claves presentes para cada dominio de la herramienta ROB1 y emitir un juicio de evaluación. La herramienta automatizada evaluó 2200 estudios, sus resultados fueron comparados con el juicio establecido por revisores humanos para estos mismos artículos. Toda la información fue obtenida de artículos incluidos en la base de datos de Cochrane Database of Systematic Reviews, la cual incluye estudios con diferentes tópicos clínicos.

En un estudio posterior, Marshall et al. (31) afinaron el modelo de aprendizaje de máquinas hacia una versión no supervisada y lo denominaron RobotReviewer. En este segundo estudio, la metodología fue similar evaluando 12808 ensayos clínicos que incluían como comparador la evaluación del riesgo de sesgos realizada por revisores humanos clínicos presentes en Cochrane Database of Systematic Reviews. Este segundo estudio reportó para el RobotReviewer una exactitud del 71% para establecer riesgo de sesgo general. (Ver Tabla 4)

El estudio realizado por Hirt et al. (62), incorporó ensayos clínicos realizados en el campo de enfermería en general. Para la evaluación automatizada del riesgo de sesgo, se utilizó el modelo RobotReviewer y se comparó con la evaluación del riesgo de sesgo realizada por humanos previamente e informada en revisiones sistemáticas Cochrane. Para esta evaluación, se incluyeron 190 ensayos clínicos controlados (98% fueron

incluidos en revisiones Cochrane). El estudio evaluó desenlaces de validez y concordancia para la herramienta automatizada. Este estudio identificó la concordancia como moderada entre el modelo de aprendizaje de máquinas y los revisores humanos, el índice Kappa fue superior al 40% en los dominios de generación de secuencia aleatoria, asignación oculta, y ocultamiento del desenlace. El área bajo la curva para el modelo automatizado se reportó por dominios, siendo de 0,75 (IC 95% 0,66-0,84) para el dominio de generación de secuencia aleatoria, 0,80 (IC 95% 0,73-0,83) para el dominio de ocultamiento de la asignación, 0,69 (IC 95% 0,52-0,78) para enmascaramiento del personal y participantes, y 0,53 (IC 95% 0,40-0,66) para ocultamiento del desenlace.

De manera similar, el estudio propuesto por Gates et al. (63) contó con RobotReviewer como herramienta automatizada. Los ensayos clínicos que tenían como muestra incluyeron como población adultos y niños e intervenciones para condiciones como bronquiolitis, migraña, diabetes, deshidratación por gastroenteritis y dolor por fractura de cadera. La evaluación del riesgo de sesgo de manera convencional fue realizada prospectivamente de

forma independiente por dos revisores, y los desacuerdos se resolvieron mediante consenso o un tercero proporcionó el arbitraje. Los desenlaces de validez y concordancia fueron medidos en este estudio. El estudio reportó que la precisión de la herramienta automatizada para establecer el riesgo general del sesgo fue 90,6% (rango 67,4- 100%) y una concordancia relativamente baja [Kappa de 0,34 (IC 95% 0,25 – 0,44)].

Finalmente, Armijo-Olivo et al. (65) en su estudio evaluaron la validez y la concordancia de RobotReview en una muestra de estudios clínicos de terapia física incluidos en revisiones sistemáticas publicadas en la base Cochrane Database of Systematic Reviews, comparando con el juicio de sesgo de manera convencional. Para esto, dos revisores capacitados y de manera independiente evaluaron los riesgos de sesgo utilizando la herramienta ROB1 de Cochrane prospectivamente, la puntuación final para cada dominio se basó en el consenso entre los dos revisores. Se incluyeron 392 estudios que evaluaron el impacto de la terapia física en condiciones musculoesqueléticas y cardiorrespiratorias principalmente. En este estudio se documentó un rendimiento bajo para clasificar el riesgo de sesgo en general con una sensibilidad de 0%, especificidad 98% y un nivel de concordancia no significativo [Kappa=−0.0139 (IC 95%=−0.029, 0.001)]. er

Tabla 3. Características de los estudios incluidos y sus resultados de concordancia

Referencia	Metodología	# de EC evaluados	Intervención	Comparador	Desenlaces/Resultados	
Hirt J, et al. (62)	Estudio diagnóstico mixto	190	Análisis de texto mediante el software RobotReviewer	Evaluación de revisores humanos informadas en las revisiones Cochrane	Concordancia y exactitud por dominios: ● Generación de secuencia aleatoria K= 0.52 (IC 95%= 0.36, 0.68) ● Ocultamiento de la asignación K= 0.60 (IC 95%=0.48, 0.72)	● Enmascaramiento del personal y participantes K=0.43 (IC 95%=0.14, 0.72) ● Ocultamiento del desenlace K= 0.04 (IC 95%=−1.14, 0.22)
Gates, et al. (63)	Estudio diagnóstico prospectivo	1180	Análisis de texto mediante el software RobotReviewer	Evaluación por 2 revisores humanos	1. Concordancia estimación global del riesgo: K= 0.34 (IC 95%=0.25, 0.44) 2. Concordancia por dominio ● Generación de secuencia aleatoria K= 0.48 (IC 95%=0.43, 0.53) ● Ocultamiento de la asignación K= 0.45(IC 95%=0.40, 0.51)	● Enmascaramiento del personal y participantes K= 0.42 (IC 95%=0.36, 0.47) ● Ocultamiento del desenlace K= 0.10 (IC 95%=0.06, 0.14)
Armijo-Olivo, et al. (65)	Estudio diagnóstico prospectivo	393	Análisis de texto mediante el software RobotReviewer	Evaluación por 6 revisores humanos	1. Concordancia general: Para clasificar bajo de riesgo de sesgo K= −0.0139 (IC 95%=−0.029, 0.001) 2. Concordancia por dominios: • Generación de secuencia aleatoria. Para detectar bajo riesgo de sesgo: K= 0.6082 (IC 95%=0.561, 0.636) ● Ocultamiento de la asignación Para detectar bajo riesgo de sesgo K= 0.4084 (IC 95%=0.305, 0.512)	● Enmascaramiento del personal y participantes Para detectar bajo riesgo de sesgo: K= 0.1207 (IC 95%=0.078, 0.158) ● Ocultamiento del desenlace Para detectar bajo riesgo de sesgo: K= 0.2319 (IC 95%=0.126, 0.337)

EC= Ensayo clínico, K= Kappa de Cohen, IC 95%= Intervalo de confianza al 95%.

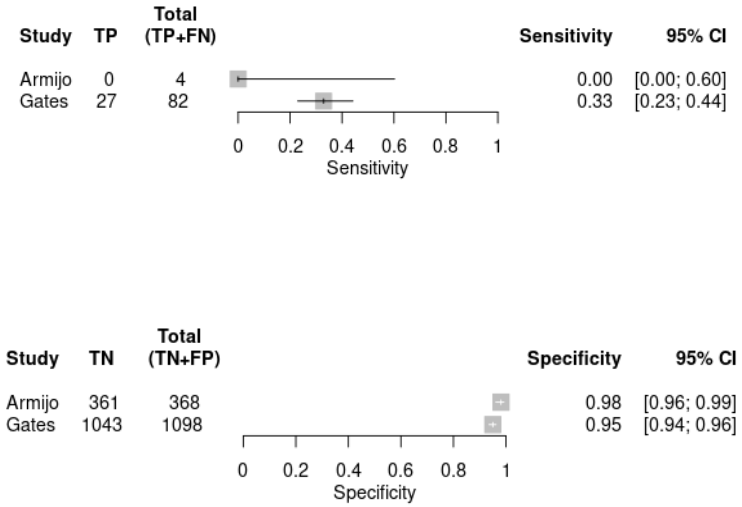
Tabla 4. Características de los estudios incluidos con resultados únicamente para validez global

Referencia	Metodología	# de EC evaluados	Intervención	Comparador	Desenlaces/Resultados	
Marshall, et al. (64)	Estudio diagnóstico mixto	2200	Modelo basado en aprendizaje de máquinas semiautomatizado	Evaluación de revisores humanos informadas en las revisiones Cochrane	• Generación de secuencia aleatoria S= 0.43 VPP= 0.68 • asignación oculta S= 0.43 VPP= 0.68 • Enmascaramiento del personal y participantes S= 0.43 VPP= 0.68	• Ocultamiento de desenlaces S= 0.43 VPP= 0.68 • Datos de resultados incompletos S=0.16 VPP=0.44 • Reporte selectivo S=0.11 VPP=0.04
Marshall IJ, et al. (31)	Estudio diagnóstico mixto	12808	Análisis de texto mediante el software RobotReviewer	Evaluación de revisores humanos informadas en las revisiones Cochrane	1. Validez general: Exactitud= 71.0% 2. Validez por dominios: • Generación de secuencia aleatoria Exactitud= 73.9% • Ocultamiento de la asignación Exactitud= 74.0%	• Enmascaramiento de participantes y personal Exactitud= 73.0% • Ocultamiento del desenlace Exactitud= 61.5%

EC= Ensayo clínico, S= sensibilidad

La validez de la herramienta automatizada fue evaluada en los estudios publicados por Hirt (62), Gates (63), Armijo (65). La Figura 2 resume la sensibilidad y especificidad para la evaluación del riesgo de sesgo en general. Esta estimación fue reportada en únicamente para los estudios de Armijo y Gates. En el estudio de Armijo, la sensibilidad fue reportada como 0%, mientras que la especificidad 98%. Los resultados de Gates fueron similares con una sensibilidad del 33% y especificidad del 95%.

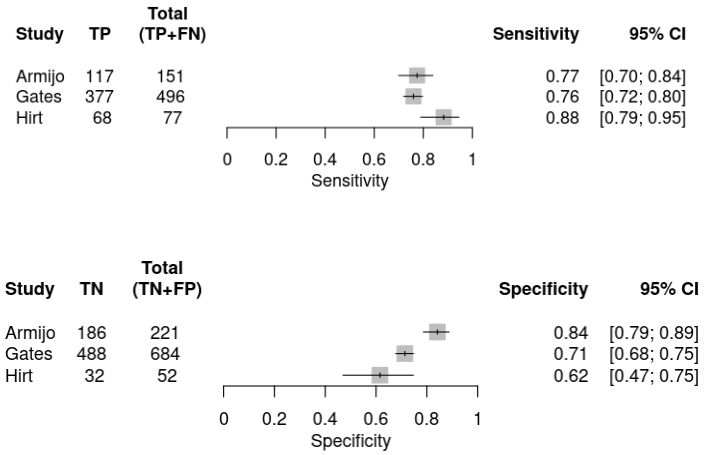
Figura 2. Estimación general de la validez de la evaluación del riesgo de sesgo para la herramienta de aprendizaje automatizado en los estudios incluidos.



TP: True positive, FP: Falso positive, FN: Falso negative, TN: True negativo

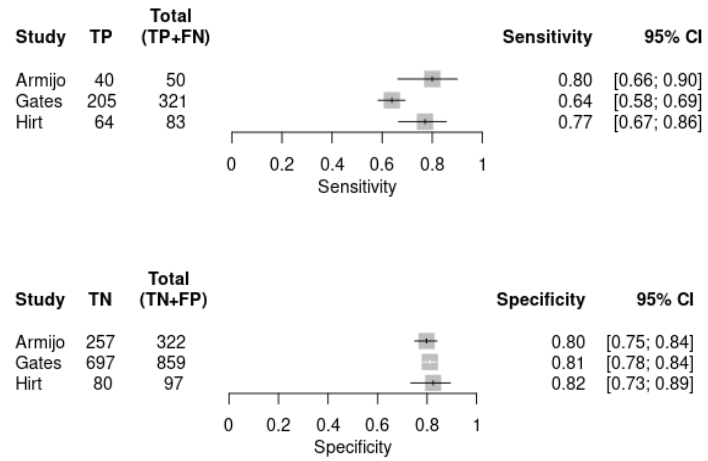
Los estimados para la validez de las herramientas automatizadas de aprendizaje de máquinas para el dominio de generación de secuencia aleatoria y asignación oculta reportados en los estudios incluidos se presentan en las Figuras 3 y 4. En relación con el dominio de generación de secuencia aleatoria la sensibilidad varió entre el 77%- 88% y la especificidad entre 62% al 84%. Para el dominio de asignación oculta la sensibilidad varió entre el 77%- 80% y la especificidad entre 80% al 82%.

Figura 3. Estimación de la validez para el dominio generación de secuencia aleatoria, de la evaluación del riesgo de sesgo para la herramienta de aprendizaje automatizado en los estudios incluidos



TP: True positive, FP: Falso positive, FN: Falso negative, TN: True negativo

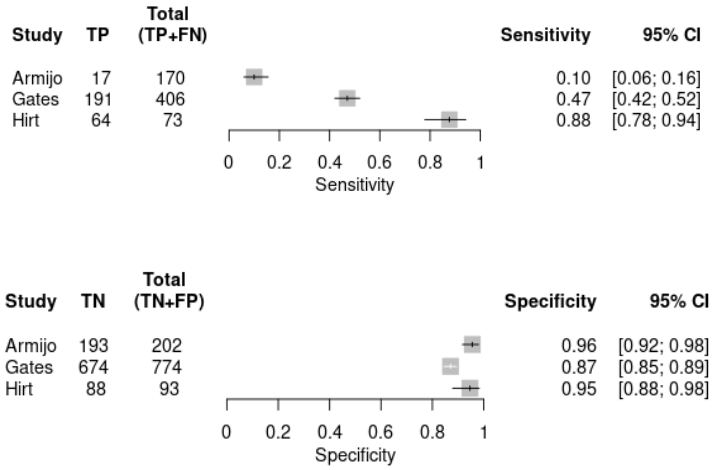
Figura 4. Estimación de la validez para el dominio asignación oculta, de la evaluación del riesgo de sesgo para la herramienta de aprendizaje automatizado en los estudios incluidos.



TP: True positive, FP: Falso positive, FN: Falso negative, TN: True negativo

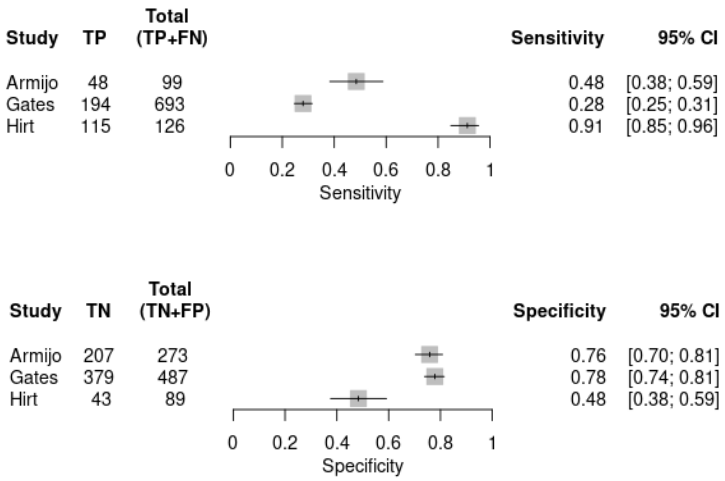
Para el dominio de enmascaramiento del personal y participantes la sensibilidad y especificidad se reportó en los diferentes estudios entre el 10 al 88% y el 87% al 96%, respectivamente (ver Figura 5). El dominio de ocultamiento de desenlace registró una alta variabilidad entre los estudios para la estimación de sensibilidad y especificidad, encontrándose entre 28 a 91% para establecer sensibilidad y 48 al 78% para la especificidad (Ver Figura 6).

Figura 5. Estimación de la validez para el dominio de enmascaramiento del personal y los participantes, de la evaluación del riesgo de sesgo para la herramienta de aprendizaje automatizado en los estudios incluidos.



TP: True positive, FP: Falso positive, FN: Falso negative, TN: True negativo

Figura 6. Estimación de la validez para el dominio de ocultamiento del desenlace, de la evaluación del riesgo de sesgo para la herramienta de aprendizaje automatizado en los estudios incluidos.



TP: True positive, FP: Falso positive, FN: Falso negative, TN: True negativo

Los estudios incluidos en la síntesis permitieron establecer las razones de verosimilitud (LR + y LR -) para cada dominio evaluado. Para el dominio de generación de secuencia aleatoria, el LR + tuvo valores de 2,65 a 4,89 y LR - 0,19 a 0,27. En el dominio de asignación oculta se identificaron valores de LR + de 3,39 a 4,40, LR - de 0,25 a 0,45, mientras que en el dominio de enmascaramiento del personal y los participantes se registraron valores de LR + desde 2,24 a 16,31 y LR - desde 0,13 a 0,94. En el dominio de ocultamiento de desenlaces el LR + se encontró desde 1,26 a 2,01 y el LR - desde 0,18 a

0,93. Los estudios publicados por Armijo (65) y Gates (63) describen la estimación global del riesgo de sesgo, donde se identifica una amplia variabilidad. Esta alta variabilidad refleja la imprecisión para la evaluación del riesgo de sesgo. La Tabla 5 describe detalladamente los hallazgos.

Tabla 5. Estimación de razones de verosimilitud del modelo de aprendizaje de máquinas para la estimación del riesgo de sesgo para cada dominio de ROB de cada estudio incluido

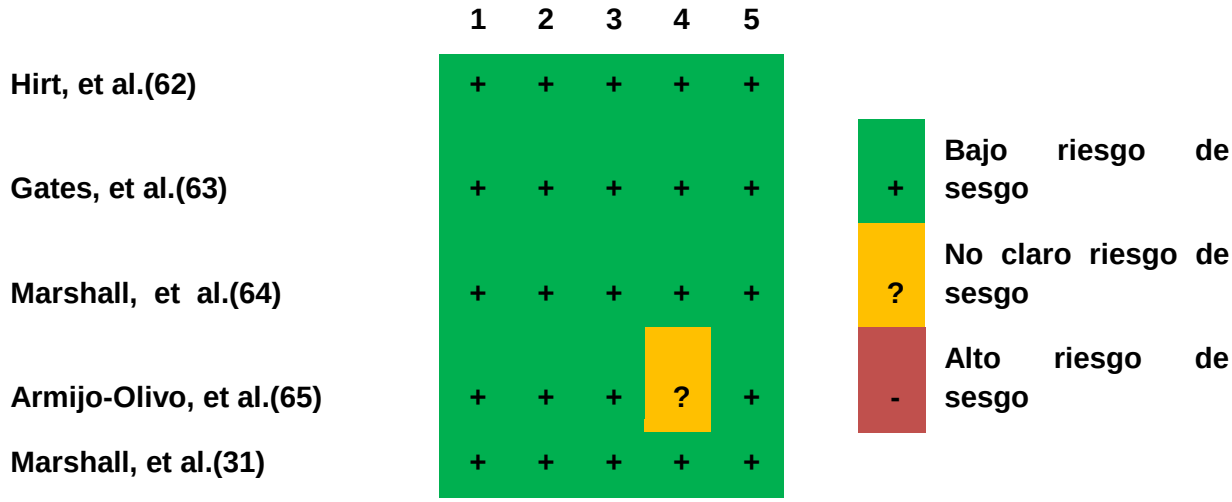
		Generación de secuencia aleatoria	Asignación oculta	Enmascaramiento del personal y participantes	Ocultamiento de desenlaces	Estimación global del riesgo de sesgo
Armijo, et al.(65)	LR +	4.89 (IC 95% 3.57, 6.71)	3.96 (IC 95% 3.06, 5.13)	2.24 (IC 95% 1.03, 4.90)	2.01 (IC 95% 1.50, 2.69)	0.00 (IC 95% 0.00 - 0.00)
	LR –	0.27 (95%IC 0.20, 0.36)	0.25 (IC 95% 0.14, 0.44)	0.94 (IC 95% 0.89, 1.00)	0.68 (IC 95% 0.55, 0.83)	1.02 (1.00, 1.03)
Gates, et al.(63)	LR +	2.65(IC 95% 2.33, 3.02)	3.39 (IC 95% 2.88, 3.98)	3.64 (IC 95% 2.95, 4.49)	1.26 (IC 95% 1.03, 1.55)	6.57 (IC 95% 4.40, 9.83)
	LR –	0.34 (95%IC 0.29, 0.40)	0.45 (IC 95% 0.38, 0.52)	0.61 (CI 95% 0.55, 0.67)	0.93 (IC 95% 0.87, 0.99)	0.71 (IC 95% 0.61, 0.82)
Hirt, et al. (62)	LR +	3.52 (95%IC 1.95, 6.34)	4.40 (IC 95% 2.81, 6.88)	16.31 (IC 95% 6.92, 38.42)	1.77 (IC 95% 1.43, 2.17)	--
	LR –	0.29 (95%IC 0.19, 0.44)	0.28 (IC 95% 0.19, 0.42)	0.13 (IC 95% 0.07, 0.24)	0.18 (IC 95% 0.10, 0.33)	--

LR + = Likelihood ratio positive, LR- = Likelihood ratio negative.

Riesgo de sesgo de los estudios incluidos en la síntesis cualitativa

La Figura 7 describe los hallazgos para la evaluación del riesgo de sesgos en los 5 estudios incluidos. Se identificó en general un bajo riesgo de sesgo. En el estudio de Armijo-Olivo, et al. (62) no hubo claridad en la evaluación ciega de los resultados.

Figura 7. Evaluación del riesgo de sesgo de los estudios



1. ¿Existió una comparación con una prueba de referencia adecuada?

2. ¿Incluyó la muestra un espectro adecuado de pacientes?

3. ¿Existe una adecuada descripción de la prueba?

4. ¿Hubo evaluación "ciega" de los resultados?

5. ¿La decisión de realizar el patrón de oro fue independiente del resultado de la prueba problema?

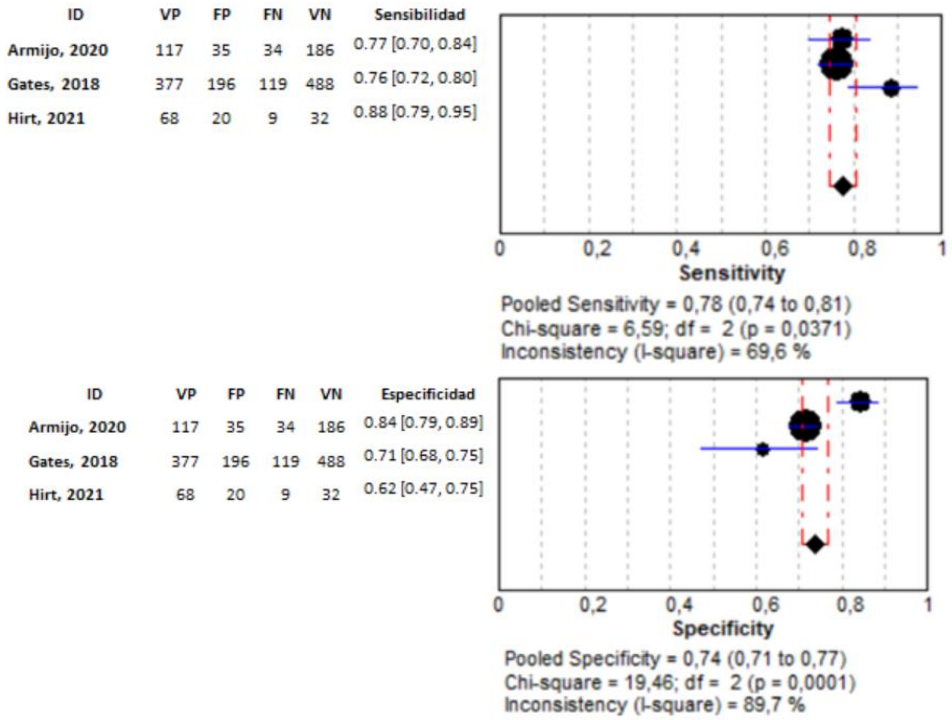
Metaanálisis

La síntesis de los estudios incluidos permitió identificar heterogeneidad cualitativa. Adicionalmente, solo fue posible obtener información relacionada con las estimaciones de sensibilidad y especificidad para 3 de los 5 estudios, para 4 de los dominios usados en la herramienta ROB 1. Debido al número limitado de estudios la síntesis cuantitativa se limitó al análisis de los diagramas de bosque y el diagrama ROC, que evidenciaron heterogeneidad estadística para las estimaciones de sensibilidad y especificidad para cada uno de los dominios evaluados. Consecuentemente, no se realizó síntesis ya que el número reducido de estudios identificados así como su alta heterogeneidad no permitió establecer el

análisis mediante un modelo bivariado o el modelo HSROC, los cuales actualmente son recomendado por la colaboración Cochrane para la realización de metanálisis de estudios de pruebas diagnósticas.

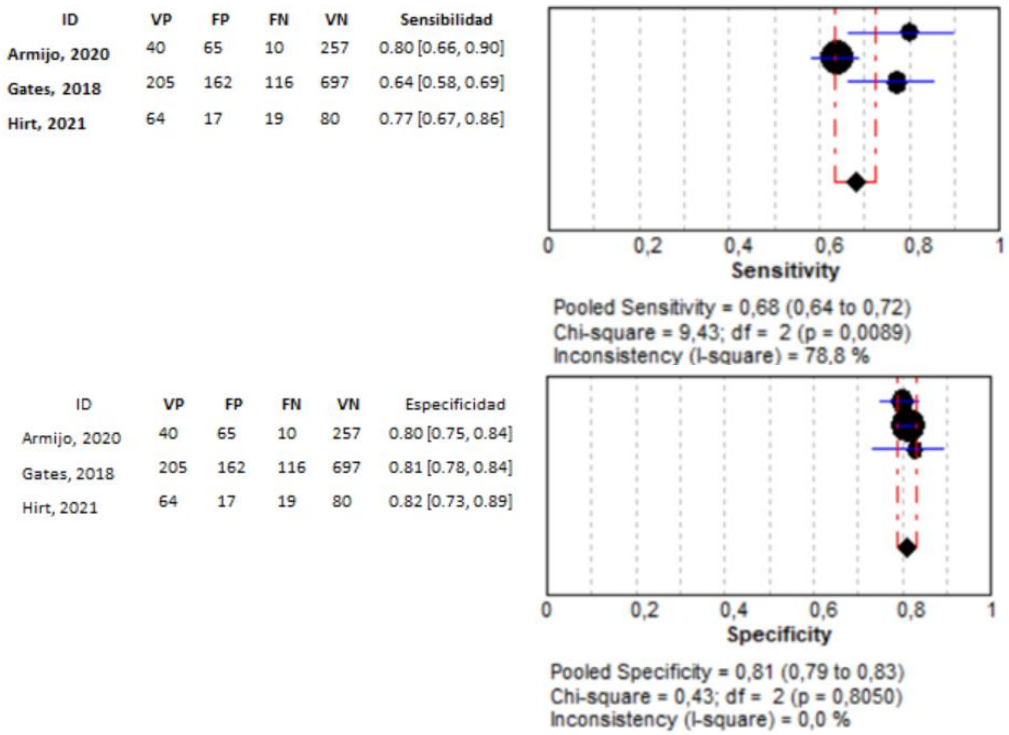
La Figura 8 presenta el estimador agrupado para la sensibilidad y especificidad para la determinación del riesgo de sesgo de la herramienta automatizada en el dominio de generación de la secuencia aleatoria. Para este dominio, la medida estimada para la sensibilidad fue de 78% (IC 95% de 74% al 81%), y para la especificidad de 74% (IC 95%= 71% al 77%).

Figura 8. Estimación de la sensibilidad y especificidad para evaluar el riesgo de sesgo en el dominio de generación de secuencia aleatoria para la herramienta automatizada de aprendizaje de máquinas.



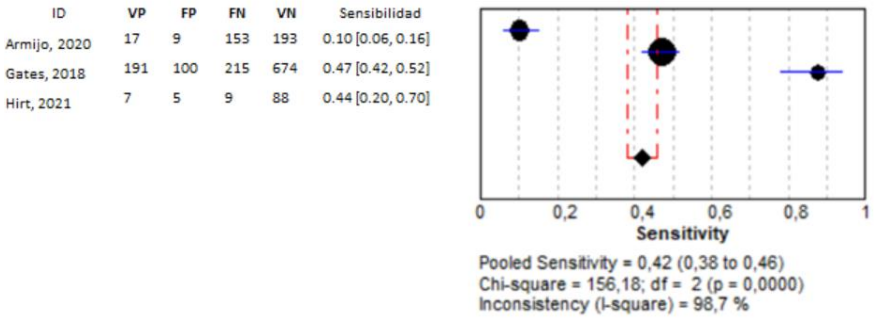
Para la evaluación del riesgo de sesgos de la herramienta automatizada en el dominio de asignación oculta, la medida estimada agrupada para la sensibilidad fue de 68% (IC 95% de 64% al 72%), y para la especificidad de 81% (IC 95%= 79% al 83%). Ver Figura 9.

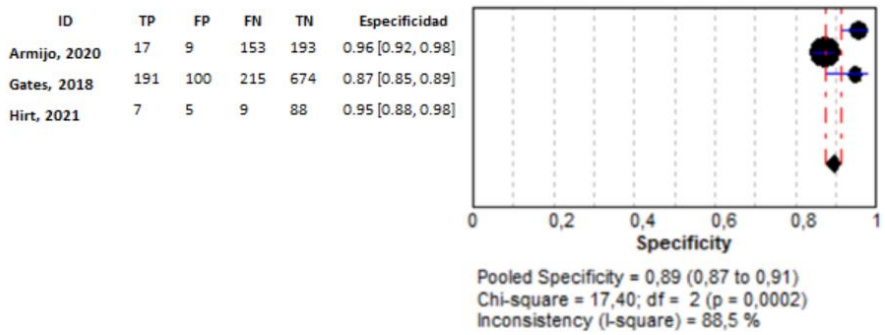
Figura 9. Estimación de la sensibilidad y especificidad para evaluar el riesgo de sesgo en el dominio asignación oculta para la herramienta automatizada de aprendizaje de máquinas.



Para la evaluación del riesgo de sesgos de la herramienta automatizada en el dominio de enmascaramiento del personal y participantes, la medida estimada agrupada para la sensibilidad fue de 42% (IC 95% de 38% al 46%), y para la especificidad de 89% (IC 95%= 87% al 91%). Ver Figura 10.

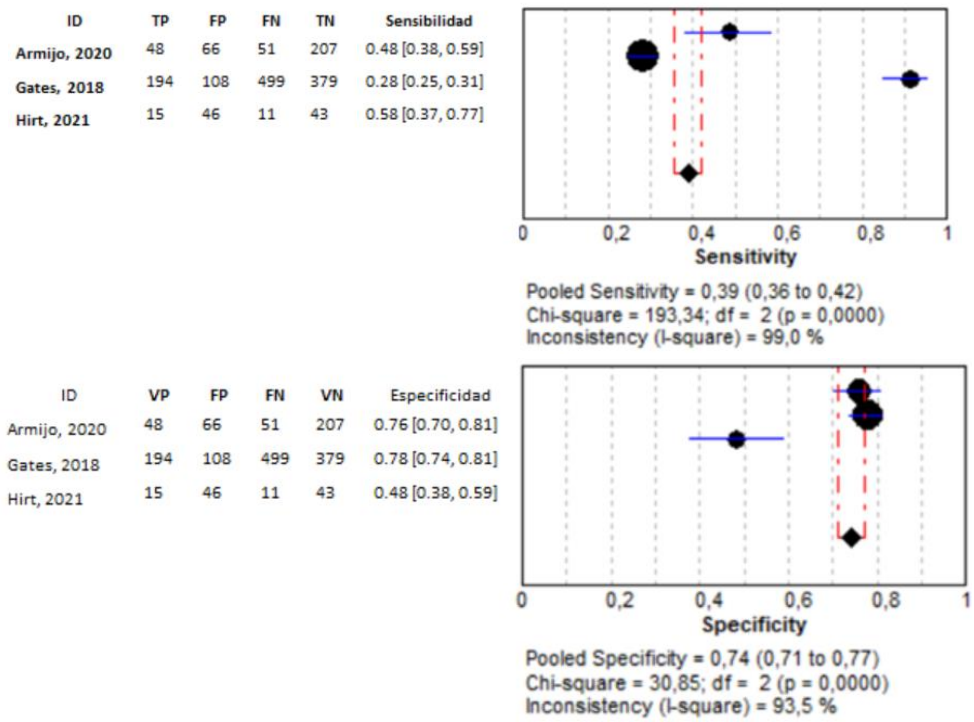
Figura 10. Estimación de la sensibilidad y especificidad para evaluar el riesgo de sesgo en el dominio enmascaramiento del personal y participantes para la herramienta automatizada de aprendizaje de máquinas.





Para la evaluación del riesgo de sesgos de la herramienta automatizada en el dominio de ocultamiento de desenlaces, la medida estimada agrupada para la sensibilidad fue de 39% (IC 95% de 36% al 42%), y para la especificidad de 74% (IC 95%= 71% al 77%). Ver Figura 11.

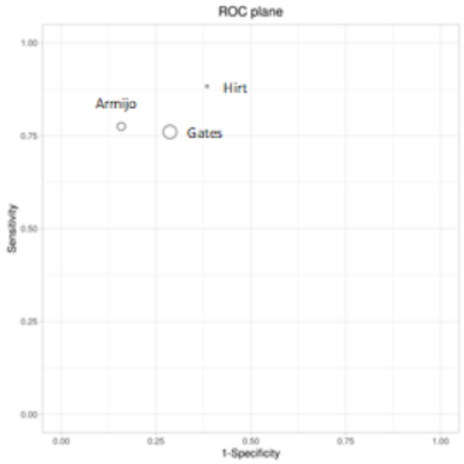
Figura 11. Estimación de la sensibilidad y especificidad para evaluar el riesgo de sesgo en el dominio ocultamiento del desenlace para la herramienta automatizada de aprendizaje de máquinas.



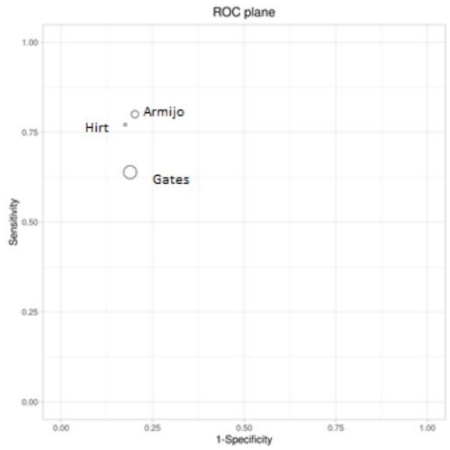
La representación gráfica del plano ROC entre la sensibilidad y 1-especificidad para cada estudio, por dominio de la evaluación del riesgo de sesgo para la herramienta automatizada de aprendizaje de máquinas se describe en la Figura 12.

Figura 12. Curvas ROC para cada dominio de la evaluación del riesgo de sesgo para la herramienta automatizada de aprendizaje de máquinas.

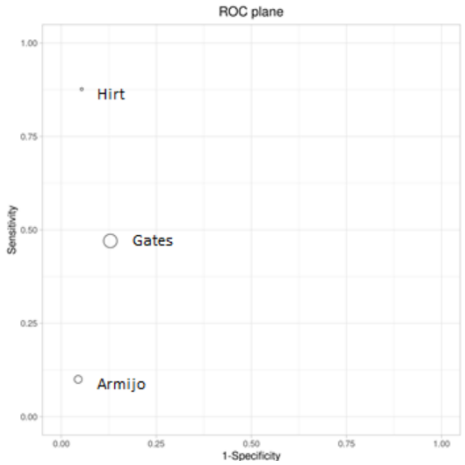
Generación secuencia aleatoria



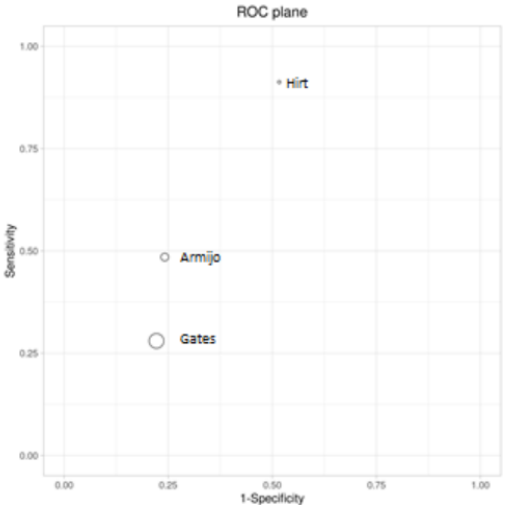
Asignación oculta



Enmascaramiento del personal y participantes



Ocultamiento de desenlaces



9. DISCUSIÓN

La evaluación del riesgo de sesgo es un paso crucial en el desarrollo de revisiones sistemáticas, dado el alcance que pueden tener éstas sobre la práctica clínica. Para su realización se ha recomendado el uso de herramientas estandarizadas; estos instrumentos son aplicados por al menos 2 investigadores de manera independiente (66,67), no obstante, es proceso es complejo y con frecuencia estos juicios pueden diferir entre los investigadores (68), por tanto, se ha propuesto como una opción la automatización de la evaluación del riesgo de sesgo de ensayos clínicos aleatorizados para optimizar el desarrollo de revisiones sistemáticas (57).

Entre los estudios incluidos en el metaanálisis, identificamos una especificidad agrupada superior a 73% del modelo de aprendizaje de máquinas para la evaluación de riesgo de sesgo en los dominios de generación de secuencia aleatoria y asignación oculta, mientras que la sensibilidad agrupada fue apenas superior al 68% para estos mismos dominios. En contraste, en los dominios como enmascaramiento del personal y participantes, y ocultamiento del desenlace, la sensibilidad agrupada se estimó en 39% y 42% respectivamente, y la especificidad agrupada con valores superiores a 73% en ambos dominios.

El análisis cualitativo de los hallazgos permitió identificar estimadores consistentes de validez de la herramienta de aprendizaje de máquinas principalmente para establecer el riesgo de sesgo en los dominios de generación de secuencia aleatoria y asignación oculta. Por el contrario, se identificó alto grado de variabilidad al estimar la validez al evaluar los dominios de ocultamiento de desenlace y enmascaramiento de los participantes y personal, siendo demostrado no solo al observar la variabilidad de la sensibilidad y especificidad, sino también esta observación es reforzada al identificar el grado en que varían los valores obtenidos de razones de verosimilitud, a pesar de estos últimos ser menos susceptibles al efecto umbral. Esto puede ser explicado por el funcionamiento de la herramienta automatizada, el cual se basa en la identificación y extracción de las palabras u oraciones claves para evaluar cada dominio dentro de cada estudio, por lo tanto, este depende de la calidad del reporte de los ensayos clínicos evaluados. Pese a la estandarización proporcionada por los lineamientos CONSORT, estudios previos han documentado un reporte deficiente para algunos de los aspectos necesarios para obtener la información relacionada con esos dominios, lo que podría resultar particularmente en términos menos explícitos para evaluar el riesgo de sesgo relacionado con el enmascaramiento del personal y participantes, y el ocultamiento de los desenlaces (69, 70, 71).

Los hallazgos en relación al desenlace concordancia permite establecer en general un nivel insignificante a bajo entre los diferentes dominios de la evaluación del riesgo de sesgo. Se observaron valores discretamente mejores en los dominios de generación de secuencia aleatoria y asignación oculta, sin que estos valores sobrepasan un nivel mediano de concordancia entre los revisores humanos y el aprendizaje de máquinas. Estos datos son concordantes con los hallazgos en sensibilidad y especificidad para estos dominios donde tuvieron mejores resultados respecto a los otros dominios evaluados. Los hallazgos revelan la complejidad técnica que puede representar la evaluación del riesgo de sesgo de ensayos clínicos, no sólo para la herramienta que se está probando, sino también para los seres humanos, donde a pesar de considerarse el *“gold standard”,* per-sé se han identificado

grandes dificultades para lograr alcanzar niveles aceptables de concordancia para la evaluación del riesgo de sesgo utilizando la herramienta ROB 1 (38).

El modelo de aprendizaje de máquinas que se identificó con mayor frecuencia en nuestra búsqueda fue RobotReviewer, el cual tiene la capacidad de identificar y extraer del texto la oración de cada dominio del instrumento de Cochrane ROB 1 (62–65). Los estudios realizados por Marshall et al. (31,64) fueron los primeros en establecer el rendimiento del aprendizaje de máquinas usando esta herramienta automatizada en comparación con la estrategia convencional de revisión por humanos, demostrando un rendimiento moderado, el cual con el paso del tiempo y la optimización de la herramienta ha demostrado una mejora significativa, lo que se refleja en la capacidad discriminatoria reportada por los estudios publicados por Hirt (62) y Gates (63), posteriormente.

Esta herramienta ha tenido una adecuada percepción entre sus usuarios, Soboczynski et al., realizaron una encuesta utilizando la escala de usabilidad del sistema (System usability scale- (SUS) scores) para establecer su utilidad, registrando altas puntuaciones sobre su favorabilidad entre los usuarios y adicionalmente, se documentó un menor tiempo para establecer el juicio de riesgo de sesgo mediante esta técnica respecto al método usual por revisiones humanos, siendo este hallazgo significativo (72).

Esta tecnología ya se ha incorporado en la producción de síntesis de evidencia, tanto en por grupos Cochrane como por otros grupos, Goldkuhle et al., del grupo Cochrane Haematological Malignancies Group realizó una revisión sobre Nivolumab para pacientes adultos con linfoma de Hodgkin (73), Bilal J et al., evaluó la eficacia y seguridad de los inhibidores de la interleuquina 1 para el manejo de la artritis psoriásica (74) y Aali G et al., utilizó está herramienta para evaluar el riesgo de sesgo de los estudios incluidos en su revisión “Post-stroke fatigue: a scoping review”, resaltando la utilidad de Robotreviewer en términos de ahorro de tiempo y otros recursos (75).

A pesar de la adecuada percepción y recibimiento de esta herramienta por parte de los usuarios, resaltamos que el modelo utiliza el instrumento ROB 1 el cual ha mostrado debilidades que podría afectar la confiabilidad para establecer el juicio de riesgo de sesgo por parte de los revisores humanos (34,35), por tanto, se considera que estas falencias del instrumento podrían trasladarse a la evaluación del riesgo de sesgo realizado por los modelos de aprendizaje de máquinas.

Es conocido que existe una versión “mejorada” del instrumento de evaluación de riesgo de sesgo denominado ROB 2 que ha sido introducida recientemente (39), no obstante, en nuestra búsqueda sistemática no identificamos trabajos que incluyeran tal instrumento como soporte para el modelo de aprendizaje de máquinas, así mismo, las evaluaciones iniciales a este instrumento sugieren una baja concordancia interevaluadores para aplicar el juicio de riesgo de sesgo de cada uno de los dominios evaluados entre revisores humanos (76). Estos hallazgos pueden estar relacionados con la reciente introducción de este instrumento y podría en corto plazo convertirse en el estándar para la evaluación del riesgo de sesgo, en la medida que los revisores adquieran experiencia en su uso. Por el momento, se desconoce el alcance que podría tener este instrumento para la evaluación del riesgo de sesgo de ensayos clínicos y su incorporación dentro de modelos de aprendizaje de máquinas.

Esta revisión sistemática tiene varias fortalezas. En primer lugar, es la primera revisión sistemática que sintetiza la información disponible sobre el uso del aprendizaje de máquinas para la evaluación del riesgo de sesgo de ensayos clínicos, permitiendo establecer su potencial de aplicación en la realización de revisiones sistemáticas. En segundo lugar, se realizó una búsqueda exhaustiva en diferentes bases de datos, ejecutada por un equipo interdisciplinario con experiencia y capacitación (1 bibliotecólogo, 1 estudiante de maestría y 1 auxiliar de investigación). En tercer lugar, los artículos incluidos en la síntesis cualitativa de este estudio fueron evaluados con bajo riesgo de sesgo, lo cual da validez a sus estimaciones identificadas e incrementa la confianza de nuestras conclusiones. Finalmente, a pesar de la presencia de heterogeneidad cualitativa y estadística entre los artículos incluidos en la revisión sistemática, y a un número limitado de estudios, se realizó como ejercicio de síntesis cuantitativa de los resultados, obteniendo estimadores agrupados y sus intervalos de confianza para los desenlaces de validez.

Las debilidades de esta revisión incluyen la posible presencia de sesgos de publicación dado que la búsqueda se limitó a los idiomas inglés y español únicamente, así como la búsqueda de literatura gris limitada a la base de datos google scholar. Por otro lado, un bajo número de estudios pudieron ser incluidos en nuestra revisión, lo cual refleja que aún existe una limitada evidencia de ausencia de información en torno al uso de aprendizaje de máquinas para la evaluación del riesgo de sesgo de ensayos clínicos que permitan a los investigadores, revisores, o metodólogos determinar el escenario de su aplicabilidad.

10. CONCLUSIONES

El aprendizaje de máquinas podría ser una herramienta de apoyo para la evaluación de riesgo de sesgo de ensayos clínicos, principalmente en el proceso de descartar aquellos estudios con alto riesgo de sesgos. La herramienta automatizada de aprendizaje de máquinas podría ser utilizada para descartar estudios con alto riesgo de sesgo en ensayos clínicos en el primer cribado durante la realización de una revisión sistemática, de esta manera podría reducir la carga de trabajo sobre los revisores humanos.

Por el momento, la evidencia sugiere que la evaluación por humanos sigue siendo indispensable, y que incorporar las herramientas de aprendizaje de máquinas no sustituye dicha evaluación.

Es deseable mejorar el rendimiento de estas herramientas para implementar su uso en el desarrollo de revisiones sistemáticas, por lo tanto, se hacen necesarios más estudios que aborden la incorporación del aprendizaje de máquinas en la automatización en este paso de la producción de síntesis de evidencia para establecer la validez de dicha prueba con mayor precisión y poder hacer una recomendación fuerte a la comunidad científica.

11. REFERENCIAS

1. NIH. National Library of Medicine. [Internet]. Pubmed.gov. [citado 15 de abril de 2022]. Disponible en: https://pubmed.ncbi.nlm.nih.gov/
2. Chalmers I. Addressing uncertainties about the effects of treatments offered to NHS patients: whose responsibility? J R Soc Med. 1 de octubre de 2007;100(10):440-1.
3. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. The Lancet. julio de 2017;390(10092):415-23.
4. Manterola C, Astudillo P, Arias E, Claros N. Revisiones sistemáticas de la literatura. Qué se debe saber acerca de ellas. Cir Esp. marzo de 2013;91(3):149-55.
5. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol. diciembre de 2011;11(1):15.
6. Muka T, Glisic M, Milic J, Verhoog S, Bohlius J, Bramer W, et al. A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research. Eur J Epidemiol. enero de 2020;35(1):49-60.
7. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction—the why, what, when, and how. J Clin Epidemiol. noviembre de 2017;91:23-30.
8. 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. noviembre de 2017;91:31-7.
9. Adam NR, Wieder R, Ghosh D. Data science, learning, and applications to biomedical and health sciences: Data science, learning, and applications to biomedical and health sciences. Ann N Y Acad Sci. enero de 2017;1387(1):5-11.
10. Ristevski B, Chen M. Big Data Analytics in Medicine and Healthcare. J Integr Bioinforma [Internet]. 25 de septiembre de 2018 [citado 15 de abril de 2022];15(3). Disponible en: https://www.degruyter.com/document/doi/10.1515/jib-2017-0030/html
11. Altman R. Artificial intelligence (AI) systems for interpreting complex medical datasets. Clin Pharmacol Ther. mayo de 2017;101(5):585-6.
12. Cheadle C, Cao H, Kalinin A, Hodgkinson J. Advanced literature analysis in a Big Data world: Advanced literature analysis. Ann N Y Acad Sci. enero de 2017;1387(1):25-33.
13. Massalha S, Clarkin O, Thornhill R, Wells G, Chow BJW. Decision Support Tools, Systems, and Artificial Intelligence in Cardiac Imaging. Can J Cardiol. julio de 2018;34(7):827-38.
14. Alami H, Lehoux P, Auclair Y, de Guise M, Gagnon MP, Shaw J, et al. Artificial Intelligence and Health Technology Assessment: Anticipating a New Level of Complexity. J Med Internet Res. 7 de julio de 2020;22(7):e17707.
15. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. enero de 2019;25(1):44-56.
16. van Dinter R, Tekinerdogan B, Catal C. Automation of systematic literature reviews: A systematic literature review. Inf Softw Technol. agosto de 2021;136:106589.
17. Harrer S, Shah P, Antony B, Hu J. Artificial Intelligence for Clinical Trial Design. Trends Pharmacol Sci. agosto de 2019;40(8):577-91.
18. Qin X, Liu J, Wang Y, Liu Y, Deng K, Ma Y, et al. Natural language processing was effective in assisting rapid title and abstract screening when updating systematic reviews. J Clin Epidemiol. mayo de 2021;133:121-9.
19. Ambalavanan AK, Devarakonda MV. Using the contextual language model BERT for multi-criteria classification of scientific articles. J Biomed Inform. diciembre de 2020;112:103578.
20. Lerner I, Créquit P, Ravaud P, Atal I. Automatic screening using word embeddings achieved high sensitivity and workload reduction for updating living network meta-

analyses. J Clin Epidemiol. abril de 2019;108:86-94.
21. Paynter RA, Featherstone R, Stoeger E, Fiordalisi C, Voisin C, Adam GP. A prospective comparison of evidence synthesis search strategies developed with and without text-mining tools. J Clin Epidemiol. noviembre de 2021;139:350-60.
22. Thomas J, McNaught J, Ananiadou S. Applications of text mining within systematic reviews. Res Synth Methods. marzo de 2011;2(1):1-14.
23. Paynter R, Bañez LL, Erinoff E, Lege-Matsuura J, Potter S. Commentary on EPC methods: an exploration of the use of text-mining software in systematic reviews. J Clin Epidemiol. abril de 2017;84:33-6.
24. Rathbone J, Hoffmann T, Glasziou P. Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers. Syst Rev. diciembre de 2015;4(1):80.
25. On behalf of the founding members of the ICASR group, Beller E, Clark J, Tsafnat G, Adams C, Diehl H, et al. Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR). Syst Rev. diciembre de 2018;7(1):77.
26. Tsou AY, Treadwell JR, Erinoff E, Schoelles K. Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer. Syst Rev. diciembre de 2020;9(1):73.
27. Popoff E, Besada M, Jansen JP, Cope S, Kanters S. Aligning text mining and machine learning algorithms with best practices for study selection in systematic literature reviews. Syst Rev. diciembre de 2020;9(1):293.
28. Chai KEK, Lines RLJ, Gucciardi DF, Ng L. Research Screener: a machine learning tool to semi-automate abstract screening for systematic reviews. Syst Rev. diciembre de 2021;10(1):93.
29. Pradhan R, Hoaglin DC, Cornell M, Liu W, Wang V, Yu H. Automatic extraction of quantitative data from ClinicalTrials.gov to conduct meta-analyses. J Clin Epidemiol. enero de 2019;105:92-100.
30. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. Syst Rev. diciembre de 2019;8(1):163, s13643-019-1074-9.
31. Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. J Am Med Inform Assoc. 1 de enero de 2016;23(1):193-201.
32. Millard LA, Flach PA, Higgins JP. Machine learning to assist risk-of-bias assessments in systematic reviews. Int J Epidemiol. febrero de 2016;45(1):266-77.
33. Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 18 de octubre de 2011;343(oct18 2):d5928-d5928.
34. Savović J, Weeks L, Sterne JA, Turner L, Altman DG, Moher D, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. Syst Rev. diciembre de 2014;3(1):37.
35. Puljak L, Ramic I, Arriola Naharro C, Brezova J, Lin YC, Surdila AA, et al. Cochrane risk of bias tool was used inadequately in the majority of non-Cochrane systematic reviews. J Clin Epidemiol. julio de 2020;123:114-9.
36. Kotz D, West R. Key concepts in clinical epidemiology: addressing and reporting sources of bias in randomized controlled trials. J Clin Epidemiol. septiembre de 2021;S0895435621003152.
37. Reveiz L, Chapman E, Asial S, Munoz S, Bonfill X, Alonso-Coello P. Risk of bias of randomized trials over time. J Clin Epidemiol. septiembre de 2015;68(9):1036-45.
38. Hartling L, Hamm MP, Milne A, Vandermeer B, Santaguida PL, Ansari M, et al. Testing the Risk of Bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. J Clin Epidemiol. septiembre de 2013;66(9):973-81.

39. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 28 de agosto de 2019;l4898.
40. Higgins J, Savović J, Page M, Elbers R, Sterne J. Chapter 8: Assessing risk of bias in a randomized trial. En: Cochrane Handbook for Systematic Reviews of Interventions version 63 [Internet]. 2022. Disponible en: www.training.cochrane.org/handbook
41. Zabor EC, Kaizer AM, Hobbs BP. Randomized Controlled Trials. Chest. julio de 2020;158(1):S79-87.
42. Siepmann T, Spieth PM, Kubasch AS, Penzlin AI, Illigens BMW, Barlinn K. Randomized controlled trials – a matter of design. Neuropsychiatr Dis Treat. junio de 2016;1341.
43. Backmann M. What's in a gold standard? In defence of randomised controlled trials. Med Health Care Philos. diciembre de 2017;20(4):513-23.
44. Boutron I, Page MJ, Higgins J, Altman DG, Lundh A, Hróbjartsson A. Chapter 7: Considering bias and conflicts of interest among the included studies. En: Cochrane Handbook for Systematic Reviews of Interventions version 63 [Internet]. Cochrane; 2022. Disponible en: www.training.cochrane.org/handbook.
45. Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 15 de marzo de 2008;336(7644):601-5.
46. Tripepi G, Jager KJ, Dekker FW, Wanner C, Zoccali C. Bias in clinical research. Kidney Int. enero de 2008;73(2):148-53.
47. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection Bias and Information Bias in Clinical Research. Nephron Clin Pract. 2010;115(2):c94-9.
48. Rivas-Ruiz F, Pérez-Vicente S, González-Ramírez AR. Bias in clinical epidemiological study designs. Allergol Immunopathol (Madr). enero de 2013;41(1):54-9.
49. Savović J, Turner RM, Mawdsley D, Jones HE, Beynon R, Higgins JPT, et al. Association Between Risk-of-Bias Assessments and Results of Randomized Trials in Cochrane Reviews: The ROBES Meta-Epidemiologic Study. Am J Epidemiol. 1 de mayo de 2018;187(5):1113-22.
50. Tovey D. Making sense of a systematic review. InnovAiT Educ Inspir Gen Pract. enero de 2021;14(1):45-51.
51. Delgado-Rodríguez M, Sillero-Arenas M. Systematic review and meta-analysis. Med Intensiva. octubre de 2018;42(7):444-53.
52. Pollock A, Berge E. How to do a systematic review. Int J Stroke. febrero de 2018;13(2):138-56.
53. Chandler J, Cumpston M, Thomas J, Higgins J, Deeks J, Clarke M. Chapter I: Introduction. En: Cochrane Handbook for Systematic Reviews of Interventions version 63 [Internet]. Cochrane; 2022. Disponible en: www.training.cochrane.org/handbook
54. Hamilton AJ, Strauss AT, Martinez DA, Hinson JS, Levin S, Lin G, et al. Machine learning and artificial intelligence: applications in healthcare epidemiology. Antimicrob Steward Healthc Epidemiol. 2021;1(1):e28.
55. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is Machine Learning? A Primer for the Epidemiologist. Am J Epidemiol. 21 de octubre de 2019;kwz189.
56. Alanazi A. Using machine learning for healthcare challenges and opportunities. Inform Med Unlocked. 2022;30:100924.
57. Khalil H, Ameen D, Zarnegar A. Tools to support the automation of systematic reviews: a scoping review. J Clin Epidemiol. abril de 2022;144:22-42.
58. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. Syst Rev. diciembre de 2014;3(1):74.
59. Cowie K, Rahmatullah A, Hardy N, Holub K, Kallmes K. Web-Based Software Tools for Systematic Literature Review in Medicine: Systematic Search and Feature Analysis. JMIR Med Inform. 2 de mayo de 2022;10(5):e33219.
60. 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. octubre de 2021;138:80-94.
61. Clark J, Glasziou P, Del Mar C, Bannach-Brown A, Stehlik P, Scott AM. A full systematic review was completed in 2 weeks using automation tools: a case study. J Clin Epidemiol. mayo de 2020;121:81-90.
62. Hirt J, Meichlinger J, Schumacher P, Mueller G. Agreement in Risk of Bias Assessment Between RobotReviewer and Human Reviewers: An Evaluation Study on Randomised Controlled Trials in Nursing-Related Cochrane Reviews. J Nurs Scholarsh. marzo de 2021;53(2):246-54.
63. Gates A, Vandermeer B, Hartling L. Technology-assisted risk of bias assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool. J Clin Epidemiol. abril de 2018;96:54-62.
64. Marshall IJ, Kuiper J, Wallace BC. Automating Risk of Bias Assessment for Clinical Trials. IEEE J Biomed Health Inform. julio de 2015;19(4):1406-12.
65. Armijo-Olivo S, Craig R, Campbell S. Comparing machine and human reviewers to evaluate the risk of bias in randomized controlled trials. Res Synth Methods. mayo de 2020;11(3):484-93.
66. Viswanathan M, Patnode CD, Berkman ND, Bass EB, Chang S, Hartling L, et al. Recommendations for assessing the risk of bias in systematic reviews of health-care interventions. J Clin Epidemiol. mayo de 2018;97:26-34.
67. Flemyng E, Dwan K, Moore TH, Page MJ, Higgins JP. Risk of Bias 2 in Cochrane Reviews: a phased approach for the introduction of new methodology. Cochrane Editorial Unit, editor. Cochrane Database Syst Rev [Internet]. 19 de noviembre de 2020 [citado 16 de abril de 2022]; Disponible en: http://doi.wiley.com/10.1002/14651858.ED000148
68. Bertizzolo L, Bossuyt P, Atal I, Ravaud P, Dechartres A. Disagreements in risk of bias assessment for randomised controlled trials included in more than one Cochrane systematic reviews: a research on research study using cross-sectional design. BMJ Open. abril de 2019;9(4):e028382.
69. Li J-Y, Zhang Y-F, Smith GS, Xue C-J, Luo Y-N, Chen W-H, et al. Quality of Reporting of Randomized Clinical Trials in Tai Chi Interventions—A Systematic Review. Evidence-Based Complementary and Alternative Medicine [Internet]. 2011 [cited 2019 Dec 14];2011:1–10. Available from: https://www.hindawi.com/journals/ecam/2011/383245/
70. Du J, Zhang Y, Dong Y, Duan J, Bai H, Wang J, et al. Reporting quality of randomized, controlled trials evaluating immunotherapy in lung cancer. Thoracic Cancer. 2021 Aug 25;12(20):2732–9.
71. McGrath M, Chen C, Braga LH, Farrokhyar F. Quality of reporting for pilot randomized controlled trials in the pediatric urology literature–A systematic review. Journal of Pediatric Urology. 2021 Sep;
72. Soboczenski F, Trikalinos TA, Kuiper J, Bias RG, Wallace BC, Marshall IJ. Machine learning to help researchers evaluate biases in clinical trials: a prospective, randomized user study. BMC Med Inform Decis Mak. diciembre de 2019;19(1):96.
73. Goldkuhle M, Dimaki M, Gartlehner G, Monsef I, Dahm P, Glossmann JP, et al. Nivolumab for adults with Hodgkin's lymphoma (a rapid review using the software RobotReviewer). Cochrane Haematological Malignancies Group, editor. Cochrane Database Syst Rev [Internet]. 12 de julio de 2018 [citado 24 de mayo de 2022];2018(7). Disponible en: https://doi.wiley.com/10.1002/14651858.CD012556.pub2
74. Bilal J, Riaz IB, Kamal MU, Elyan M, Sudano D, Khan MA. A Systematic Review and Meta-analysis of Efficacy and Safety of Novel Interleukin Inhibitors in the Management of Psoriatic Arthritis: JCR J Clin Rheumatol. enero de 2018;24(1):6-13.
75. Aali G, Drummond A, das Nair R, Shokraneh F. Post-stroke fatigue: a scoping review. F1000Research. 25 de agosto de 2020;9:242.
76. Minozzi S, Cinquini M, Gianola S, Gonzalez-Lorenzo M, Banzi R. The revised Cochrane risk of bias tool for randomized trials (RoB 2) showed low interrater reliability and challenges in its application. J Clin Epidemiol. octubre de 2020;126:37-44.

12. ANEXO 1

Estudios identificados al aplicar las ecuaciones de búsquedas en las diferentes bases de datos.

En la tabla a continuación se describen todos los estudios identificados en las diferentes bases de datos al aplicar las ecuaciones de búsquedas preestablecidas.

	Fuente	Título	Resumen	autores	publicador T2	Afiliación	Lugar de publicación	Año	Disponible en digital
1	Scopus	What's in a Name? Nothing, (in PubMed) if You Change It!	Most physician-scientists are unaware that PubMed, the search tool used to access MEDLINEindexed peer-reviewed articles, cannot link citations if they undergo a name change during their academic careers. For example, author L.N.K. had attempted to find a publication by a colleague who was a rising star in her field and in doing so realized that the colleague must have become divorced. She eventually found the PubMed citation after trying several variations of the author's name; the older elusive paper was indexed under the author's married hyphenated surname, whereas her newer publications were under her maiden name. A social media check confirmed the divorce; a PubMed search confirmed the disconnect between her 2 names. This colleague has approximately 20 publications indexed under her hyphenated name (including first-author articles in JAMA and The New England Journal of Medicine) and more than 90 indexed by her maiden name. Because PubMed searches cannot link alternate versions of an author's name, one could miss almost 1 in 5 of her manuscripts. This issue became more personal when we realized that the daughter of author S.N.H., who had changed her surname when she married during graduate school, had twice as many publications under her maiden name than under her married name. If one were unaware of the name change, one would inaccurately assume that her academic career began in 2018 instead of 2012	Kransdorf L.N. ; Hayes S.N	Mayo Clinic Proceedings	University of California ; Mayo Clinic	Estados Unidos	2021	Clinical Key
2	Scopus	Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI	In recent years a substantial literature has emerged concerning bias, discrimination, and fairness in artificial intelligence (AI) and machine learning. Connecting this work to existing legal non-discrimination frameworks is essential to create tools and methods that are practically useful across divergent legal regimes. While much work has been undertaken from an American legal perspective, comparatively little has mapped the effects and requirements of EU law. This Article addresses this critical gap between legal, technical, and organisational notions of algorithmic fairness. Through analysis of EU non-discrimination law and jurisprudence of the European Court of Justice (ECJ) and national courts, we identify a critical incompatibility between European notions of discrimination and existing work on algorithmic and automated fairness. A clear gap exists between statistical measures of fairness as embedded in myriad fairness toolkits and governance mechanisms and the context-sensitive, often intuitive and ambiguous discrimination metrics and evidential requirements used by the ECJ; we refer to this approach as "contextual equality." This Article makes three contributions. First, we review the evidential requirements to bring a claim under EU non-discrimination law. Due to the disparate nature of algorithmic and human discrimination, the EU's current requirements are too contextual, reliant on intuition, and open to	Wachter S. ; Mittelstadt B. ; Russell C.	Computer Law and Security Review	University of Oxford, Harvard University, British Library, University of Surrey	Estados Unidos, Reino Unido	2021	Disponible en Science Direct

			judicial interpretation to be automated. Many of the concepts fundamental to bringing a claim, such as the composition of the disadvantaged and advantaged group, the severity and type of harm suffered, and requirements for the relevance and admissibility of evidence, require normative or political choices to be made by the judiciary on a case-by-case basis. We show that automating fairness or non-discrimination in Europe may be impossible because the law, by design, does not provide a static or homogenous framework suited to testing for discrimination in AI systems. Second, we show how the legal protection offered by non-discrimination law is challenged when AI, not humans, discriminate. Humans discriminate due to negative attitudes (e.g. stereotypes, prejudice) and unintentional biases (e.g. organisational practices or internalised stereotypes) which can act as a signal to victims that discrimination has occurred. Equivalent signalling mechanisms and agency do not exist in algorithmic systems. Compared to traditional forms of discrimination, automated discrimination is more abstract and unintuitive, subtle, intangible, and difficult to detect. The increasing use of algorithms disrupts traditional legal remedies and procedures for detection, investigation, prevention, and correction of discrimination which have predominantly relied upon intuition. Consistent assessment procedures that define a common standard for statistical evidence to detect and assess prima facie automated discrimination are urgently needed to support judges, regulators, system controllers and developers, and claimants. Finally, we examine how existing work on fairness in machine learning lines up with procedures for assessing cases under EU non-discrimination law. A 'gold standard' for assessment of prima facie discrimination has been advanced by the European Court of Justice but not yet translated into standard assessment procedures for automated discrimination. We propose 'conditional demographic disparity' (CDD) as a standard baseline statistical measurement that aligns with the Court's 'gold standard'. Establishing a standard set of statistical evidence for automated discrimination cases can help ensure consistent procedures for assessment, but not judicial interpretation, of cases involving AI and automated systems. Through this proposal for procedural regularity in the identification and assessment of automated discrimination, we clarify how to build considerations of fairness into automated systems as far as possible while still respecting and enabling the contextual approach to judicial interpretation practiced under EU non-discrimination law. © 2021 The Authors						
3	Scopus	Using artificial intelligence to improve COVID-19 rapid diagnostic test result interpretation	Serological rapid diagnostic tests (RDTs) are widely used across pathologies, often providing users a simple, binary result (positive or negative) in as little as 5 to 20 min. Since the beginning of the COVID-19 pandemic, new RDTs for identifying SARS-CoV-2 have rapidly proliferated. However, these seemingly easy-to-read tests can be highly subjective, and interpretations of the visible "bands" of	Mendels D.-A., Dortet L., Emeraud C., Oueslati	Proceedings of the National Academy of Sciences of the United States of		Francia	2021	Si

			color that appear (or not) in a test window may vary between users, test models, and brands. We developed and evaluated the accuracy/performance of a smartphone application (xRCovid) that uses machine learning to classify SARS-CoV-2 serological RDT results and reduce reading ambiguities. Across 11 COVID-19 RDT models, the app yielded 99.3% precision compared to reading by eye. Using the app replaces the uncertainty from visual RDT interpretation with a smaller uncertainty of the image classifier, thereby increasing confidence of clinicians and laboratory staff when using RDTs, and creating opportunities for patient self-testing. © 2021 National Academy of Sciences. All rights reserved.	S., Girlich D., Ronat J.-B., Bernabeu S., Bahi S., Atkinson G.J.H., Naas T.	America				
4	Scopus	Agreement in Risk of Bias Assessment Between RobotReviewer and Human Reviewers: An Evaluation Study on Randomised Controlled Trials in Nursing-Related Cochrane Reviews	Purpose: RobotReviewer is a machine learning system for semi-automated assistance in risk of bias assessment. The tools's performance in randomized controlled trials (RCTs) in the field of nursing remains unknown. We aimed therefore to evaluate the agreement in risk of bias assessment between RobotReviewer and human reviewers. Design: Evaluation study using a retrospective diagnostic design. Methods: We used RobotReviewer as the index test and human reviewers' risk of bias assessment reported in Cochrane reviews as the reference test. A convenience sample of electronically available English-language full texts of RCTs included in Cochrane reviews with nurs* in the title were eligible for inclusion. In this context, we assessed random sequence generation, allocation concealment, and blinding (personnel or participants and assessors) corresponding to Cochrane risk of bias version 2011. Two independent research teams performed and double-checked data extraction and analysis. We calculated sensitivity, specificity, receiver operating characteristic (ROC) curve, the area under the ROC curve, predictive values, observed percentage of agreement, and Cohen's kappa (including confidence intervals, if applicable). Findings: The selection process yielded 190 RCTs published between 1958 and 2016 in 23 Cochrane reviews published between 2000 and 2018. Missing assessments of risk of bias domains in Cochrane reviews or RobotReviewer yielded varying sample sizes per risk of bias domain. Sensitivity ranged from 0.44 to 0.88 and specificity from 0.48 to 0.95. Positive predictive value was highest for allocation concealment (0.79) and lowest for blinding assessors (0.25). Cohen's kappa was moderate for randomization (0.52), allocation concealment (0.60), and for blinding of personnel/patients (0.43). Blinding of outcome assessors had only slight agreement (0.04). Conclusions: This is the first evaluation of risk of bias assessment by RobotReviewer in RCTs included in nursing-related Cochrane reviews. It yielded a moderate degree of agreement with human reviewers for randomization and allocation concealment, and an adequate sensitivity for detecting low risk of selection bias. Clinical Relevance: Based on our results, using the RobotReviewer for risk of bias assessment in RCTs can be supportive in some risk of bias domains. However, human reviewers should supervise the	Hirt J., Meichlinger J., Schumacher P., Mueller G.	Journal of Nursing Scholarship	University for Applied Sciences FHS, Martin Luther University Halle-Wittenberg, IMC University of Applied Sciences Krems, UMIT–Private University for Health Sciences	Suiza, Alemania, Austria	2021	Si

			semi-automated assessment process. © 2021 The Authors. Journal of Nursing Scholarship published by Wiley Periodicals LLC on behalf of Sigma Theta Tau International.						
5	Scopus	Machine learning's limitations in avoiding automation of bias	The use of predictive systems has become wider with the development of related computational methods, and the evolution of the sciences in which these methods are applied Solon and Selbst (Calif L REV 104: 671–732, 2016) and Pedreschi et al. (2007). The referred methods include machine learning techniques, face and/or voice recognition, temperature mapping, and other, within the artificial intelligence domain. These techniques are being applied to solve problems in socially and politically sensitive areas such as crime prevention and justice management, crowd management, and emotion analysis, just to mention a few. However, dissimilar predictions can be found nowadays as the result of the application of these methods resulting in misclassification, for example for the case of conviction risk assessment Office of Probation and Pretrial Services (2011) or decision-making process when designing public policies Lange (2015). The goal of this paper is to identify current gaps on fairness achievement within the context of predictive systems in artificial intelligence by analyzing available academic and scientific literature up to 2020. To achieve this goal, we have gathered available materials at the Web of Science and Scopus from last 5 years and analyzed the different proposed methods and their results in relation to the bias as an emergent issue in the Artificial Intelligence field of study. Our tentative conclusions indicate that machine learning has some intrinsic limitations which are leading to automate the bias when designing predictive algorithms. Consequently, other methods should be explored; or we should redefine the way current machine learning approaches are being used when building decision making/decision support systems for crucial institutions of our political systems such as the judicial system, just to mention one. © 2020, Springer-Verlag London Ltd., part of Springer Nature.	Varona D., Lizama-Mue Y.a, Suárez J.L.	AI and Society	Western University	Canada	2021	Disponible en Springer Link
6	Scopus	Humans, Automation, and Workflow: Seeking Balance in Service of Art	Since the advent of 'workflow' during the Industrial Revolution, companies have been continually seeking to increase efficiency while reducing extraneous tasks or human interaction, including through automation. The media and entertainment (M&E) industry has followed a similar trajectory, accelerated most recently by the transition to digital. While automation is inherent to modern	Chaet S., List P., McIntire R.	SMPTE Motion Imaging Journal			2021	No

			workflows, M&E relies on creativity and the subjectivity of art and therefore requires human intervention. Mitigating issues of algorithm bias, over-reliance on technology, and limited or flawed data inputs creating flawed outputs is vital in promoting healthy collaboration that allows elevated creativity. © 2002 Society of Motion Picture and Television Engineers, Inc.						
7	Scopus	Regimes of recognition on algorithmic media	This article examines ways in which people are seen, recognised, and made to matter by social media platforms. Drawing on Louise Amoore's notion of 'regimes of recognition', I argue that social media platforms can be conceptualised as increasingly powerful arbiters of recognisability, determining the conditions of possibility of how people are seen and come to matter. Through an analysis of Twitter's saliency detection algorithm, which automatically crops images uploaded to the platform, the article highlights how social media platforms participate in producing novel modes of recognisability, that is, conditions by which people are rendered visible and invisible within or by the platform. Moreover, the article highlights how regimes of recognition on algorithmic media shape people's parameters of attention and perception more generally through what I call the automatic production of 'consistent' lines of sight. Ultimately, the article seeks to highlight how the notion of recognition is increasingly arbitrated in and through algorithmic media and how this is fraught with political issues and tension. As such, there is an ongoing need to critically examine the power of social media to render people visible and invisible. © The Author(s) 2021.	Jacobsen B.N.	New Media and Society	Durham University	Reino Unido	2021	Si
8	Scopus	ObjectivAIze: Measuring Performance and Biases in Augmented Business Decision Systems	Business process management organizes flows of information and decisions in large organizations. These systems now integrate algorithmic decision aids leveraging machine learning: each time a stakeholder needs to make a decision, such as a purchase, a quote, or hiring someone, the software leverages the inputs and outcomes of similar past decisions to provide guidance, as a recommendation. If the confidence is high, the process may be automated. Otherwise, it may still help provide consistency in the decisions. Yet, we may question how these aids affect task performance. Can we measure an improvement? Can hidden biases influence decision makers negatively? What is the impact of various presentation options? To address those issues, we propose metrics of performance, automation bias and resistance. We validated those measures with an online study. Our aim is to instrument those systems to secure their benefits. In a first experiment, we study effective collaboration. Faced with a decision, subjects alone have a success rate of 72%; Aided by a recommender that has a 75% success rate, their success rate reaches 76%. The human-system collaboration had thus a greater success rate than each taken alone. However, we noted a complacency/authority bias that degraded the quality of decisions by 5% when the recommender was wrong. This suggests that any lingering algorithmic bias may be amplified by decision aids. In a	Baudel T., Verbockhaven M., Cousergue V., Roy G., Laarach R.	Lecture Notes in Computer Science	IBM, ENSAE, Université Paris-Dauphine, Mines ParisTech, ENSAI, Telecom ParisTech	Francia	2021	Disponible en Springer Link

			second experiment, we evaluated the effectiveness of 5 presentation variants in reducing complacency bias. We found that optional presentation increases subjects' resistance to wrong recommendations. We intend to leverage these findings to guide the design of human-algorithm collaboration in financial compliance alert filtering. © 2021, IFIP International Federation for Information Processing.						
9	Scopus	Deep automation bias: How to tackle a wicked problem of ai?	The increasing use of AI in different societal contexts intensified the debate on risks, ethical problems and bias. Accordingly, promising research activities focus on debiasing to strengthen fairness, accountability and transparency in machine learning. There is, though, a tendency to fix societal and ethical issues with technical solutions that may cause additional, wicked problems. Alternative analytical approaches are thus needed to avoid this and to comprehend how societal and ethical issues occur in AI systems. Despite various forms of bias, ultimately, risks result from eventual rule conflicts between the AI system behavior due to feature complexity and user practices with limited options for scrutiny. Hence, although different forms of bias can occur, automation is their common ground. The paper highlights the role of automation and explains why deep automation bias (DAB) is a metarisk of AI. Based on former work it elaborates the main influencing factors and develops a heuristic model for assessing DAB-related risks in AI systems. This model aims at raising problem awareness and training on the sociotechnical risks resulting from AI-based automation and contributes to improving the general explicability of AI systems beyond technical issues. © 2021 by the author. Licensee MDPI, Basel, Switzerland.	Strauß S.	Big Data and Cognitive Computing	Austrian Academy of Sciences	Austria	2021	Si
10	Scopus	Addressing bias: artificial intelligence in cardiovascular medicine	Artificial intelligence (AI) is providing opportunities to transform cardiovascular medicine. As the leading cause of morbidity and mortality worldwide, cardiovascular disease is prevalent across all populations, with clear benefit to operationalise clinical and biomedical data to improve workflow, quality of patient care, and lower costs. AI technologies are designed to aid accurat decisions, without the bias of external factors—a pitfall of human errors. However, AI algorithms are still subject to their own biases	Tat E., Bhatt D.L., Rabbat M.G.	The Lancet Digital Health	Columbia University Medical Center, Harvard Medical School, Loyola University Medical Center	Estados Unidos	2020	Disponible en Science Direct
11	Scopus	Bias and ethical considerations in machine learning and the automation of perioperative risk assessment	Anaesthesiologists have an ever-increasing number of tools at their disposal to assess the risk that a patient will assume when undergoing anaesthesia and surgery, and the likelihood of specific events during surgery. The earliest ver- sions of these tools were based on heuristics or on a small number of easily identifiable features; the Revised Cardiac Risk Index (RCRI) and pulse pressure variation are good examples of this type of preoperative and intraoperative assessment tool. However, as medicine and technology have evolved, extremely large amounts of healthcare data have become available as computation and data storage	O'Reilly-Shah V.N., Gentry K.R., Walters A.M., Zivot J., Anderson C.T., Tighe P.J.	British Journal of Anaesthesia	University of Washington School of Medicine, Seattle Children's Hospital, Emory University, University of Florida	Estados Unidos	2020	Disponible en Science Direct

			capabilities have grown ever cheaper						
12	Scopus	Features of human-centred algorithm design	Algorithms are pervasive, unseen influencers of decisions. Algorithmic features can fluctuate widely, depending on use, user or criteria applied. This paper considers the nascent field of human-centred algorithm design (HCAD), intersecting human-centred design and algorithmic systems. Human-centred, more-than-metric feature selection approaches, create fairer and deeper meaning. More value is created. The unique impact of this paper is to integrate feature selection within a technology HCAD strategy, for a novel, innovative HCAD approach to machine learning. This flexible and evaluative approach can support data advances with human-social nuance, designed for purpose with knowledge for data-driven decisions. The design of machine learning algorithms to the uses in which they will be employed is user-centric. This is important within environments utilising automated, semi-automated or high-performance analytics. © 2020 IEEE.	Cherrington M., Airehrour D., Lu J., Xu Q, Cameron-Brown D., Dunn I.	2020 30th International Telecommunication Networks and Applications Conference, ITNAC 2020	University of Huddersfield, Unitec Institute of Technology, Cameron Brown Corporate Advisory, University of Auckland	Reino Unido, Nueva Zelanda	2020	No
13	Scopus	Unbiased corneal tissue analysis using Gabor-domain optical coherence microscopy and machine learning for automatic segmentation of corneal endothelial cells	Significance: An accurate, automated, and unbiased cell counting procedure is needed for tissue selection for corneal transplantation. Aim: To improve accuracy and reduce bias in endothelial cell density (ECD) quantification by combining Gabor-domain optical coherence microscopy (GDOCM) for three-dimensional, wide field-of-view (1 mm2) corneal imaging and machine learning for automatic delineation of endothelial cell boundaries. Approach: Human corneas stored in viewing chambers were imaged over a wide field-of-view with GDOCM without contacting the specimens. Numerical methods were applied to compensate for the natural curvature of the cornea and produce an image of the flattened endothelium. A convolutional neural network (CNN) was trained to automatically delineate the cell boundaries using 180 manually annotated images from six corneas. Ten additional corneas were imaged with GDOCM and compared with specular microscopy (SM) to determine performance of the combined GDOCM and CNN to achieve automated endothelial counts relative to current procedural standards. Results: Cells could be imaged over a larger area with GDOCM than SM, and more cells could be delineated via automatic cell segmentation than via manual methods. ECD obtained from automatic cell segmentation of GDOCM images yielded a correlation of 0.94 ($p < 0.001$) with the manual segmentation on the same images, and correlation of 0.91 ($p < 0.001$) with the corresponding manually counted SM results. Conclusions: Automated endothelial cell counting on GDOCM images with large field of view eliminates selection bias and reduces sampling error, which both affect the gold standard of manual counting on SM images. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or	Canavesi C., Cogliati A., Hindman H.B.	Journal of Biomedical Optics	LighTopTech Corp., Eye Care Center	Estados Unidos	2020	Si

			reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.						
14	Scopus	Assessing clinicians' reliance on computational aids for acute stroke diagnosis	The rapid rise of computational aids for stroke diagnosis have led to important concerns about clinicians developing an over-dependence on technology. Other studies have assessed reliance on clinical decision support systems in fields like diabetes, but no such study exists for stroke diagnosis. In this work, we developed a high-fidelity user interface for a computational aid designed to support acute ischemic stroke diagnosis. Engaging with stroke practitioners at the UCSD Stroke Center, we conducted an experiment to determine how technology for identifying stroke symptoms may affect their diagnostic decision-making processes. By assessing how clinicians changed their video-based diagnosis of stroke when provided with data visualizations and predictions from a machine learning tool, we observed that such computational aids do in fact affect clinicians' decisions but only in cases when the aid directly supports or contradicts their prior beliefs. Future computational aids for stroke diagnosis should focus on helping clinicians solidify their decisions rather than only providing them with overly quantitative information that may impede or confuse their judgement. © 2020 Owner/Author.	Ramesh V., Nguyen A., Agrawal K., Meyer B.C., Cauwenberghs G., Weibel N.	PervasiveHealth: Pervasive Computing Technologies for Healthcare	University of California	Estados Unidos	2020	Si
15	Scopus	When machines think for us: The consequences for work and place	The relationship between technology and work, and concerns about the displacement effects of technology and the organisation of work, have a long history. The last decade has seen the proliferation of academic papers, consultancy reports and news articles about the possible effects of Artificial Intelligence (AI) on work-creating visions of both utopian and dystopian workplace futures. AI has the potential to transform the demand for labour, the nature of work and operational infrastructure by solving complex problems with high efficiency and speed. However, despite hundreds of reports and studies, AI remains an enigma, a newly emerging technology, and its rate of adoption and implications for the structure of work are still only beginning to be understood. The current anxiety about labour displacement anticipates the growth and direct use of AI. Yet, in many ways, at present AI is likely being overestimated in terms of impact. Still, an increasing body of research argues the consequences for work will be highly uneven and depend on a range of factors, including place, economic activity, business culture, education levels and gender, among others. We appraise the history and the blurry boundaries around the definitions of AI. We explore the debates around the extent of job augmentation, substitution,	Clifton J., Glasmeier A., Gray M.	Cambridge Journal of Regions, Economy and Society	Universidad de Cantabria, University of Cambridge, Massachusetts Institute of Technology,	Estados Unidos, Reino Unido, España	2020	Si

			destruction and displacement by examining the empirical basis of claims, rather than mere projections. Explorations of corporate reactions to the prospects of AI penetration, and the role of consultancies in prodding firms to embrace the technology, represent another perspective onto our inquiry. We conclude by exploring the impacts of AI changes in the quantity and quality of labour on a range of social, geographic and governmental outcomes. © 2020 The Author(s).						
16	Scopus	Investigating the importance of first impressions and explainable AI with interactive video analysis	We present research on how the perception of intelligent systems can be influenced by early experiences of machine performance, and how explainability potentially helps users develop an accurate understanding of system capabilities. Using a custom video analysis system with AI-assisted activity recognition, we studied whether presenting explanatory information for system outputs affects user perception of the system. In this experiment, some participants encountered AI weaknesses early, while others encountered the same limitations later in the study. The difference in ordering had a significant impact on user understanding of the system and the ability to detect AI strengths and weaknesses, and the addition of explanations was not enough to counteract the strong effects of early impressions. The results demonstrate the importance of first impressions with intelligent systems and motivate the need for improved methods of intervention to combat automation bias. © 2020 Owner/Author.	Nourani M., Honeycutt D.R., Block J.E., Roy C., Rahman T., Ragan E.D., Gogate V.	Conference on Human Factors in Computing Systems - Proceedings	University of Texas in Dallas, University of Florida	Estados Unidos	2020	No
17	Scopus	Fully Automated, Quality-Controlled Cardiac Analysis From CMR: Validation and Large-Scale Application to Characterize Cardiac Function	Objectives: This study sought to develop a fully automated framework for cardiac function analysis from cardiac magnetic resonance (CMR), including comprehensive quality control (QC) algorithms to detect erroneous output. Background: Analysis of cine CMR imaging using deep learning (DL) algorithms could automate ventricular function assessment. However, variable image quality, variability in phenotypes of disease, and unavoidable weaknesses in training of DL algorithms currently prevent their use in clinical practice. Methods: The framework consists of a pre-analysis DL image QC, followed by a DL algorithm for biventricular segmentation in long-axis and short-axis views, myocardial feature-tracking (FT), and a post-analysis QC to detect erroneous results. The study validated the framework in healthy subjects and cardiac patients by comparison against manual analysis (n = 100) and evaluation of the QC steps' ability to detect erroneous results (n = 700). Next, this method was used to obtain reference values for cardiac function metrics from the UK Biobank. Results: Automated analysis correlated highly with manual analysis for left and right ventricular volumes (all r > 0.95), strain (circumferential r = 0.89, longitudinal r > 0.89), and filling and ejection rates (all r ≥ 0.93). There was no significant bias for cardiac volumes and filling and ejection rates, except for right ventricular end-systolic volume (bias +1.80 ml; p = 0.01). The bias for FT strain was <1.3%. The sensitivity of detection	Ruijsink B., Puyol-Antón E., Oksuz I., Sinclair M., Bai W., Schnabel J.A., Razavi R., King A.P.	JACC: Cardiovascular Imaging	King's College London, Guy's and St Thomas' NHS Foundation Trust, Imperial College London	Reino Unido	2020	Disponible en Science Direct

			of erroneous output was 95% for volume-derived parameters and 93% for FT strain. Finally, reference values were automatically derived from 2,029 CMR exams in healthy subjects. Conclusions: The study demonstrates a DL-based framework for automated, quality-controlled characterization of cardiac function from cine CMR, without the need for direct clinician oversight. © 2020						
18	Scopus	Measuring complacency in humans interacting with autonomous agents in a multi-agent system	With advances in machine learning, autonomous agents are increasingly able to navigate uncertain operational environments, as is the case within the multi-domain operations (MDO) paradigm. When teaming with humans, autonomous agents may flexibly switch between passive bystander and active executor depending on the task requirements and the actions being taken by partners (whether human or agent). In many tasks, it is possible that a well-trained agent's performance will exceed that of a human, in part because the agent's performance is less likely to degrade over time (e.g., due to fatigue). This potential difference in performance might lead to complacency, which is a state defined by over-trust in automated systems. This paper investigates the effects of complacency in human-agent teams, where agents and humans have the same capabilities in a simulated version of the predator-prey pursuit task. We compare subjective measures of the human's predisposition to complacency and trust using various scales, and we validate their beliefs by quantifying complacency through various metrics associated with the actions taken during the task with trained agents of varying reliability levels. By evaluating the effect of complacency on performance, we can attribute a degree of variation in human performance in this task to complacency. We can then account for an individual human's complacency measure to customize their agent teammates and human-in-the-loop requirements (either to minimize or compensate for the human's complacency) to optimize team performance. © 2020 SPIE.	Rodriguez S.S., Chen J., Deep H., Lee J.J., Asher D.E., Zaroukian E.	Proceedings of SPIE - The International Society for Optical Engineering	University of Illinois at Urbana-Champaign	Estados Unidos	2020	No
19	Scopus	Addressing Bias in Artificial Intelligence in Health Care		Parikh R.B., Teeple S., Navathe A.S.	JAMA - Journal of the American Medical Association	University of Pennsylvania,	Estados Unidos	2019	No
20	Scopus	Machine Learning Ensembles and Rail Defects Prediction: Multilayer Stacking Methodology	Machine learning has taken a front seat in railway big data analysis. This is partly due to perpetual data collection and the need for automated systems to expedite maintenance decisions. A case for track defect prognosis in rail track engineering is presented in this paper. Fatigue defects are very common and are influential on rail maintenance. Understanding such defects is essential for optimized maintenance scheduling. The literature is replete with machine learning models developed for defect prediction. Because no single	Lasisi A., Attoh-Okine N.	ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering	Univ. of Delaware	Estados Unidos	2019	No

			machine learning model is guaranteed to surpass others with every kind of data, each model has its inherent deficiencies. Classifier ensembles such as bagging or boosting aggregate strengths from different models to enhance prediction. The outcome is very effective, although highly correlated. This work proposes a stacking method of combining average learners into powerful learning machines while considering memory, time, computational, structural complexities, and bias-variance trade-offs. Because of the large scale of rail infrastructure considered in this work (35,406 km), this study shows that classical Weibull analysis underestimates annual fatigue defects by at least 25% throughout rail life. The proposed stacking ensemble compensates for this shortfall by aggregating the probability predictions of diverse learners. These predictions were combined from a binary classification ensemble of 0.783 receiver operating characteristic area under curve (ROC-AUC) score with significant room for improvement in computation time and curve fitting. © 2019 American Society of Civil Engineers.						
21	Scopus	Human intervention in automated decision-making: Toward the construction of contestable systems	Concerns about "black box" machine learning algorithms have influenced why modern data protection laws and regulations on their establishment of a right to human intervention on decision-making supported by artificial intelligence. Such interventions provide data subjects with means to protect their rights, freedoms, and legitimate interests, either as a bare minimum requirement for data processing or as a central norm governing decision-aiding artificial intelligence. In this paper, I present contestability by design as an approach to two kinds of issues with current legal implementations of the right to human intervention. The first kind is the uncertainty about what kind of decision should be covered by this right: should intervention be restricted to those decisions with no human involvement, or should it be interpreted in a broader sense, encompassing all decisions that are effectively shaped by automated processing? The second class of issues ensues from practical limitations of this right to intervention: even within a clear conceptual framework, data subjects might still lack the information they need to the concrete exercise of their right, or the human intervention itself might introduce biases and limitations that result in undesirable outcomes. After discussing how those effects can be identified and measured, I then advance the thesis that proper protection of the rights of data subjects is feasible only if there are means for contesting decisions based solely on automated processing is not an afterthought, but instead a requirement at each stage of an artificial intelligence system's lifecycle. © 2019 Association for Computing Machinery.	Almada M.	Proceedings of the 17th International Conference on Artificial Intelligence and Law, ICAIL 2019	University of São Paulo and Lawgorithm	Brasil	2019	No

22	Scopus	Machine learning to help researchers evaluate biases in clinical trials: A prospective, randomized user study	Objective: Assessing risks of bias in randomized controlled trials (RCTs) is an important but laborious task when conducting systematic reviews. RobotReviewer (RR), an open-source machine learning (ML) system, semi-automates bias assessments. We conducted a user study of RobotReviewer, evaluating time saved and usability of the tool. Materials and methods: Systematic reviewers applied the Cochrane Risk of Bias tool to four randomly selected RCT articles. Reviewers judged: whether an RCT was at low, or high/unclear risk of bias for each bias domain in the Cochrane tool (Version 1); and highlighted article text justifying their decision. For a random two of the four articles, the process was semi-automated: users were provided with ML-suggested bias judgments and text highlights. Participants could amend the suggestions if necessary. We measured time taken for the task, ML suggestions, usability via the System Usability Scale (SUS) and collected qualitative feedback. Results: For 41 volunteers, semi-automation was quicker than manual assessment (mean 755 vs. 824 s; relative time 0.75, 95% CI 0.62-0.92). Reviewers accepted 301/328 (91%) of the ML Risk of Bias (RoB) judgments, and 202/328 (62%) of text highlights without change. Overall, ML suggested text highlights had a recall of 0.90 (SD 0.14) and precision of 0.87 (SD 0.21) with respect to the users' final versions. Reviewers assigned the system a mean 77.7 SUS score, corresponding to a rating between "good" and "excellent". Conclusions: Semi-automation (where humans validate machine learning suggestions) can improve the efficiency of evidence synthesis. Our system was rated highly usable, and expedited bias assessment of RCTs. © 2019 The Author(s).	Soboczenski F., Trikalinos T.A., Kuiper J., Bias R.G.d, Wallace B.C., Marshall I.J.a	BMC Medical Informatics and Decision Making	King's College London, Brown University, University of Texas at Austin	Estados Unidos, Reino Unido	2019	Disponible en Springer Link
23	Scopus	Artificial intelligence, bias and clinical safety	In medicine, artificial intelligence (AI) research is becoming increasingly focused on applying machine learning (ML) techniques to complex problems, and so allowing computers to make predictions from large amounts of patient data, by learning their own associations	Challen R., Denny J., Pitt M., Gompels L., Edwards T., Tsaneva-Atanasova K.	BMJ Quality and Safety	University of Exeter College of Engineering Mathematics and Physical Sciences, Vanderbilt University Medical Center, University of Exeter Medical School	Estados Unidos, Reino Unido	2019	Si
24	Scopus	On present use of machine learning based automation in finance	In this paper, we survey the current known applications of Machine Learning based Data Analytics and automation in finance industry. We look into the challenges involved in furthering this technology, particularly in employing more Deep Learning approaches proven for successful automation in other domains. We enumerate observations on some of the barriers faced by the industry in effectively adopting and accelerating use of AI techniques, and finally propose more areas that we believe could further benefit from	Tripathi V.	IC3K 2019 - Proceedings of the 11th International Joint Conference on Knowledge Discovery,	MIO	Estados Unidos	2019	No

			application of Machine Learning. Copyright © 2019 by SCITEPRESS - Science and Technology Publications, Lda. All rights reserved		Knowledge Engineering and Knowledge Management				
25	Scopus	Automation bias in medicine: The influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms	Introduction: Interpretation of the 12-lead Electrocardiogram (ECG) is normally assisted with an automated diagnosis (AD), which can facilitate an 'automation bias' where interpreters can be anchored. In this paper, we studied, 1) the effect of an incorrect AD on interpretation accuracy and interpreter confidence (a proxy for uncertainty), and 2) whether confidence and other interpreter features can predict interpretation accuracy using machine learning. Methods: This study analysed 9000 ECG interpretations from cardiology and non-cardiology fellows (CFs and non-CFs). One third of the ECGs involved no ADs, one third with ADs (half as incorrect) and one third had multiple ADs. Interpretations were scored and interpreter confidence was recorded for each interpretation and subsequently standardised using sigma scaling. Spearman coefficients were used for correlation analysis and C5.0 decision trees were used for predicting interpretation accuracy using basic interpreter features such as confidence, age, experience and designation. Results: Interpretation accuracies achieved by CFs and non-CFs dropped by 43.20% and 58.95% respectively when an incorrect AD was presented ($p < 0.001$). Overall correlation between scaled confidence and interpretation accuracy was higher amongst CFs. However, correlation between confidence and interpretation accuracy decreased for both groups when an incorrect AD was presented. We found that an incorrect AD disturbs the reliability of interpreter confidence in predicting accuracy. An incorrect AD has a greater effect on the confidence of non-CFs (although this is not statistically significant it is close to the threshold, $p = 0.065$). The best C5.0 decision tree achieved an accuracy rate of 64.67% ($p < 0.001$), however this is only 6.56% greater than the no-information-rate. Conclusion: Incorrect ADs reduce the interpreter's diagnostic accuracy indicating an automation bias. Non-CFs tend to agree more with the ADs in comparison to CFs, hence less expert physicians are more effected by automation bias. Incorrect ADs reduce the interpreter's confidence and also reduces the predictive power of confidence for predicting accuracy (even more so for non-CFs). Whilst a statistically significant model was developed, it is difficult to predict interpretation accuracy using machine learning on basic features such as interpreter confidence, age, reader experience and designation. © 2018 Elsevier Inc.	Bond R.R., Novotny T., Andrsova I., Koc L., Sisakova M., Finlay D., Guldenring D., McLaughlin J., Peace A., McGilligan V., Leslie S.J., Wang H.	Journal of Electrocardiology	Ulster University, University Hospital Brno and Faculty of Medicine of Masaryk University,	Reino Unido, Republicha Checa	2018	Disponible en Science Direct
26	Scopus	Computer systems fit for the legal profession?	This essay aims to contribute robust grounds to question the Susskinds' influential, consequentialist logic when it comes to the legitimacy of automation within the legal profession. It does so by questioning their minimalist understanding of the professions. If it is our commitment to moral equality that is at stake every time lawyers	Delacroix S.	Legal Ethics	University of Birmingham	Reino Unido	2018	Si

			(fail to) hail the specific vulnerability inherent in their professional relationship, the case for wholesale automation is turned on its head. One can no longer assume that, as a rule, wholesale automation is both legitimate and desirable, provided it improves the quality and accessibility of legal services (in an accountable and maximally transparent way). The assumption, instead, is firmly in favour of designing systems that better enable legal professionals to live up to their specific responsibility. The rest of the essay outlines key challenges in the design of such profession-specific, 'ethics aware' decision-support systems. © 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.						
27	Scopus	Bias detectives: The researchers striving to make algorithms fair news-feature		Courtland R.	Nature		Estados Unidos	2018	Disponible en Nature
28	Scopus	Technology-assisted risk of bias assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool	Objectives: To evaluate the reliability of RobotReviewer's risk of bias judgments. Study Design and Setting: In this prospective cross-sectional evaluation, we used RobotReviewer to assess risk of bias among 1,180 trials. We computed reliability with human reviewers using Cohen's kappa coefficient and calculated sensitivity and specificity. We investigated differences in reliability by risk of bias domain, topic, and outcome type using the chi-square test in meta-analysis. Results: Reliability (95% CI) was moderate for random sequence generation (0.48 [0.43, 0.53]), allocation concealment (0.45 [0.40, 0.51]), and blinding of participants and personnel (0.42 [0.36, 0.47]); fair for overall risk of bias (0.34 [0.25, 0.44]); and slight for blinding of outcome assessors (0.10 [0.06, 0.14]), incomplete outcome data (0.14 [0.08, 0.19]), and selective reporting (0.02 [−0.02, 0.05]). Reliability for blinding of participants and personnel ($P < 0.001$), blinding of outcome assessors ($P = 0.005$), selective reporting ($P < 0.001$), and overall risk of bias ($P < 0.001$) differed by topic. Sensitivity and specificity (95% CI) ranged from 0.20 (0.18, 0.23) to 0.76 (0.72, 0.80) and from 0.61 (0.56, 0.65) to 0.95 (0.93, 0.96), respectively. Conclusion: Risk of bias appraisal is subjective. Compared with reliability between author groups, RobotReviewer's reliability with human reviewers was similar for most domains and better for allocation concealment, blinding of participants and personnel, and overall risk of bias. © 2018 Elsevier Inc.	Gates A., Vandermeer B., Hartling L.	Journal of Clinical Epidemiology	University of Alberta	Canada	2018	Disponible en Science Direct
29	Scopus	Science of science	Identifying fundamental drivers of science and developing predictive models to capture its evolution are instrumental for the design of policies that can improve the scientific enterprise-for example, through enhanced career paths for scientists, better performance evaluation for organizations hosting research, discovery of novel effective funding vehicles, and even identification of promising regions along the scientific frontier. The science of science uses large-scale data on the production of science to search for universal and domain-specific patterns. Here, we review recent developments in this transdisciplinary field. © 2018 American Association for the	Fortunato S., Bergstrom C.T., Börner K., Evans J.A., Helbing D., Milojević	Science	Indiana University, University of Washington, University of Chicago, University of California, Central European	Estados Unidos, Suiza, Hungria, Italia, Holanda	2018	No

			Advancement of Science. All rights reserved.	S., Petersen A.M., Radicchi F., Sinatra R., Uzzi B., Vespignani A., Waltman L.		University, Northeastern University, Northwestern University, Leiden University,			
30	Scopus	Considerations for automated machine learning in clinical metabolic profiling: Altered homocysteine plasma concentration associated with metformin exposure	With the maturation of metabolomics science and proliferation of biobanks, clinical metabolic profiling is an increasingly opportunistic frontier for advancing translational clinical research. Automated Machine Learning (AutoML) approaches provide exciting opportunity to guide feature selection in agnostic metabolic profiling endeavors, where potentially thousands of independent data points must be evaluated. In previous research, AutoML using high-dimensional data of varying types has been demonstrably robust, outperforming traditional approaches. However, considerations for application in clinical metabolic profiling remain to be evaluated. Particularly, regarding the robustness of AutoML to identify and adjust for common clinical confounders. In this study, we present a focused case study regarding AutoML considerations for using the Tree-Based Optimization Tool (TPOT) in metabolic profiling of exposure to metformin in a biobank cohort. First, we propose a tandem rank-accuracy measure to guide agnostic feature selection and corresponding threshold determination in clinical metabolic profiling endeavors. Second, while AutoML, using default parameters, demonstrated potential to lack sensitivity to low-effect confounding clinical covariates, we demonstrated residual training and adjustment of metabolite features as an easily applicable approach to ensure AutoML adjustment for potential confounding characteristics. Finally, we present increased homocysteine with long-term exposure to metformin as a potentially novel, non-replicated metabolite association suggested by TPOT; an association not identified in parallel clinical metabolic profiling endeavors. While warranting independent replication, our tandem rank-accuracy measure suggests homocysteine to be the metabolite feature with largest effect, and corresponding priority for further translational clinical research. Residual training and adjustment for a potential confounding effect by BMI only slightly modified the suggested association. Increased homocysteine is thought to be associated with vitamin B12 deficiency – evaluation for potential clinical relevance is suggested. While considerations for clinical metabolic profiling are recommended, including adjustment approaches for clinical confounders, AutoML presents an exciting tool to enhance clinical metabolic profiling and advance translational research	Orlenko A., Moore J.H., Orzechowski P., Olson R.S., Cairns J., Caraballo P.J., Weinshilboum R.M., Wang L., Breitenstein M.K.	Pacific Symposium on Biocomputing	University of Pennsylvania, AGH University of Science and Technology, Mayo Clinic, University of Pennsylvania	Estados Unidos, Polonia	2018	No

			endeavors. © 2017 The Authors.						
31	Scopus	The use of automated Ki67 analysis to predict Oncotype DX risk-of-recurrence categories in early-stage breast cancer	Ki67 is a commonly used marker of cancer cell proliferation, and has significant prognostic value in breast cancer. In spite of its clinical importance, assessment of Ki67 remains a challenge, as current manual scoring methods have high inter- and intra-user variability. A major reason for this variability is selection bias, in that different observers will score different regions of the same tumor. Here, we developed an automated Ki67 scoring method that eliminates selection bias, by using whole-slide analysis to identify and score the tumor regions with the highest proliferative rates. The Ki67 indices calculated using this method were highly concordant with manual scoring by a pathologist (Pearson's r = 0.909) and between users (Pearson's r = 0.984). We assessed the clinical validity of this method by scoring Ki67 from 328 whole-slide sections of resected early-stage, hormone receptor-positive, human epidermal growth factor receptor 2-negative breast cancer. All patients had Oncotype DX testing performed (Genomic Health) and available Recurrence Scores. High Ki67 indices correlated significantly with several clinico-pathological correlates, including higher tumor grade (1 versus 3, P<0.001), higher mitotic score (1 versus 3, P<0.001), and lower Allred scores for estrogen and progesterone receptors (P = 0.002, 0.008). High Ki67 indices were also significantly correlated with higher Oncotype DX risk-of-recurrence group (low versus high, P<0.001). Ki67 index was the major contributor to a machine learning model which, when trained solely on clinico-pathological data and Ki67 scores, identified Oncotype DX high- and low-risk patients with 97% accuracy, 98% sensitivity and 80% specificity. Automated scoring of Ki67 can thus successfully address issues of consistency, reproducibility and accuracy, in a manner that integrates readily into the workflow of a pathology laboratory. Furthermore, automated Ki67 scores contribute significantly to models that predict risk of recurrence in breast cancer. © 2018 Thakur et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.	Thakur S.S., Li H., Chan A.M.Y., Tudor R., Bigras G., Morris D., Enwere E.K., Yang H.	PLoS ONE	University of Calgary, University of Alberta	Canada	2018	Si

32	Scopus	Artificial intelligence and the ethics of self-learning robots	The convergence of robotics technology with the science of artificial intelligence is rapidly enabling the development of robots that emulate a wide range of intelligent human behaviors. Recent advances in machine learning techniques have produced artificial agents that can acquire highly complex skills formerly thought to be the exclusive province of human intelligence. These developments raise a host of new ethical concerns about the responsible design, manufacture, and use of robots enabled with artificial intelligence-particularly those equipped with self-learning capacities. While the potential benefits of self-learning robots are immense, their potential dangers are equally serious. While some warn of a future where AI escapes the control of its human creators or even turns against us, this chapter focuses on other, far less cinematic risks of AI that are much nearer to hand, requiring immediate study and action by technologists, lawmakers, and other stakeholders. © Oxford University Press 2017. All rights reserved.	Vallor S., Bekey G.A.	Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence	Santa Clara University, University of Southern California	Estados Unidos	2017	No
33	Scopus	Machine learning in a graph framework for subcortical segmentation	Automated and reliable segmentation of subcortical structures from human brain magnetic resonance images is of great importance for volumetric and shape analyses in quantitative neuroimaging studies. However, poor boundary contrast and variable shape of these structures make the automated segmentation a tough task. We propose a 3D graph-based machine learning method, called LOGISMOS-RF, to segment the caudate and the putamen from brain MRI scans in a robust and accurate way. An atlas-based tissue classification and bias-field correction method is applied to the images to generate an initial segmentation for each structure. Then a 3D graph framework is utilized to construct a geometric graph for each initial segmentation. A locally trained random forest classifier is used to assign a cost to each graph node. The max-flow algorithm is applied to solve the segmentation problem. Evaluation was performed on a dataset of T1-weighted MRI's of 62 subjects, with 42 images used for training and 20 images for testing. For comparison, FreeSurfer, FSL and BRAINSCut approaches were also evaluated using the same dataset. Dice overlap coefficients and surface-to-surfaces distances between the automated segmentation and expert manual segmentations indicate the results of our method are statistically significantly more accurate than the three other methods, for both the caudate (Dice: 0.89 ± 0.03) and the putamen (0.89 ± 0.03). © 2017 SPIE.	Guo Z., Kashyap S., Sonka M., Oguz I.	Progress in Biomedical Optics and Imaging - Proceedings of SPIE	Univ. of Iowa, University of Pennsylvania	Estados Unidos	2017	No
34	Scopus	Exploiting reviewers' comment histories for sentiment analysis	Sentiment analysis is used to extract people's opinion from their online comments in order to help automated systems provide more precise recommendations. Existing sentiment analysis methods often assume that the comments of any single reviewer are independent of each other and so they do not take advantage of significant information that may be extracted from reviewers' comment histories. Using psychological findings and the theory of negativity bias, we propose a method for exploiting reviewers' comment histories to improve sentiment analysis. Furthermore, to	Basiri M.E., Ghasem-Aghaee N., Naghsh-Nilchi A.R.	Journal of Information Science	University of Isfahan	Iran	2017	No

			use more fine-grained information about the content of a review, our method predicts the overall ratings by aggregating sentence-level scores. In the proposed system, the Dempster-Shafer theory of evidence is utilized for score aggregation. The results from four large and diverse socialWeb datasets establish the superiority of our approach in comparison with the state-of-the-art machine learning techniques. In addition, the results show that the suggested method is robust to the size of training dataset. © The Author(s) 2014.						
35	Scopus	Automated segmentation and object classification of CT images: Application to in vivo molecular imaging of avian embryos	Background. Although chick embryogenesis has been studied extensively, there has been growing interest in the investigation of skeletogenesis. In addition to improved poultry health and minimized economic loss, a greater understanding of skeletal abnormalities can also have implications for human medicine. True in vivo studies require noninvasive imaging techniques such as high-resolution microCT. However, the manual analysis of acquired images is both time consuming and subjective. Methods. We have developed a system for automated image segmentation that entails object-based image analysis followed by the classification of the extracted image objects. For image segmentation, a rule set was developed using Definiens image analysis software. The classification engine was implemented using the WEKA machine learning tool. Results. Our system reduces analysis time and observer bias while maintaining high accuracy. Applying the system to the quantification of long bone growth has allowed us to present the first true in ovo data for bone length growth recorded in the same chick embryos. Conclusions. The procedures developed represent an innovative approach for the automated segmentation, classification, quantification, and visualization of microCT images. MicroCT offers the possibility of performing longitudinal studies and thereby provides unique insights into the morpho- and embryogenesis of live chick embryos. © 2013 Alexander Heidrich et al.	Heidrich A., Schmidt J., Zimmermann J.	International Journal of Biomedical Imaging	Leibniz Institute for Natural Product Research and Infection Biology	Alemania	2013	Si
36	Scopus	Automated bias shift in a constrained space for logic program synthesis	We propose a new approach to first order inductive learning using techniques borrowed from the state of the art constructive inductive ILP systems. In this respect a learning system ALPS is presented which performs a top-down iterative broadening search through the hypothesis space. ALPS uses argument selection heuristic of constructive inductive ILP systems which enables it to avoid a huge search space. It employs an automated bias adjustment procedure through a sequence of hypothesis subspaces arranged in a hierarchical lattice. Some experiments show that in benchmark logic program synthesis tasks, ALPS visits much less search space than well-known existing algorithms which perform a hill-climbing search through the hypothesis space. ALPS is also shown to be more successful in learning situations where there exists many irrelevant background predicates and where the training set comes from an unbiased source.	Chowdhury M.R., Numao M.	Transactions of the Japanese Society for Artificial Intelligence	Bangladesh Univ. Eng. and Technol., Tokyo Institute of Technology	Bangladesh, Japon	2001	Si

37	Google Academico	Automating Risk of Bias Assessment for Clinical Trials	Systematic reviews, which summarize the entirety of the evidence pertaining to a specific clinical question, have become critical for evidence-based decision making in healthcare. But such reviews have become increasingly onerous to produce due to the exponentially expanding biomedical literature base. This study proposes a step toward mitigating this problem by automating risk of bias assessment in systematic reviews, in which reviewers determine whether study results may be affected by biases (e.g., poor randomization or blinding). Conducting risk of bias assessment is an important but onerous task. We thus describe a machine learning approach to automate this assessment, using the standard Cochrane Risk of Bias Tool which assesses seven common types of bias. Training such a system would typically require a large labeled corpus, which would be prohibitively expensive to collect here. Instead, we use distant supervision, using data from the Cochrane Database of Systematic Reviews (a large repository of systematic reviews), to pseudoannotate a corpus of 2200 clinical trial reports in PDF format. We then develop a joint model which, using the full text of a clinical trial report as input, predicts the risks of bias while simultaneously extracting the text fragments supporting these assessments. This study represents a step toward automating or semiautomating extraction of data necessary for the synthesis of clinical trials.	Iain J Marshall; Joël Kuiper; Byron C Wallace	IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS			2015	Disponible en IEEE
38	Google Academico	Toward systematic review automation: a practical guide to using machine learning tools in research synthesis	Technologies and methods to speed up the production of systematic reviews by reducing the manual labour involved have recently emerged. Automation has been proposed or used to expedite most steps of the systematic review process, including search, screening, and data extraction. However, how these technologies work in practice and when (and when not) to use them is often not clear to practitioners. In this practical guide, we provide an overview of current machine learning methods that have been proposed to expedite evidence synthesis. We also offer guidance on which of these are ready for use, their strengths and weaknesses, and how a systematic review team might go about using them in practice.	Iain J. Marshall & Byron C. Wallace	Systematic Reviews			2019	Disponible en Springer Link
39	Google Academico	Evaluating Robotreviewer for Automated Risk of Bias Assessment in a Systematic Review: A Case Study	Risk of bias (RoB) assessment is an important part of a systematic review and hence the production of health technology assessment. However, it is a time consuming, subjective and labour intensive process, and disagreements on a study's RoB between reviewers are common. In response, novel software tools have emerged which aim to support this process. RobotReviewer is a free web-based machine learning system that aims to automate RoB assessments of randomised controlled trials (RCTs). RobotReviewer has been tested internally by its developers where it performed well, but to date we have not identified any published independent evaluations. We compared and evaluated RobotReviewer against the current standard for RoB assessment, defined as double, independent, human researcher assessment with disagreements resolved by a third reviewer.	Edwards M; Marshall C	Value in Health			2017	Disponible en Science Direct

40	Google Academico	Machine learning to help researchers evaluate biases in clinical trials: a prospective, randomized user study	Assessing risks of bias in randomized controlled trials (RCTs) is an important but laborious task when conducting systematic reviews. RobotReviewer (RR), an open-source machine learning (ML) system, semi-automates bias assessments. We conducted a user study of RobotReviewer, evaluating time saved and usability of the tool.	Frank Soboczenski , Thomas A. Trikalinos, Joël Kuiper, Randolph G. Bias, Byron C. Wallace and Iain J. Marshall	BMC Medical Informatics and Decision Making			2019	Si
41	Google Academico	Comparing machine and human reviewers to evaluate the risk of bias in randomized controlled trials	Evidence from new health technologies is growing, along with demands for evidence to inform policy decisions, creating challenges in completing health technology assessments (HTAs)/systematic reviews (SRs) in a timely manner. Software can decrease the time and burden by automating the process, but evidence validating such software is limited. We tested the accuracy of RobotReviewer, a semi-autonomous risk of bias (RoB) assessment tool, and its agreement with human reviewers.	Susan Armijo-Olivo, Rodger Craig, Sandy Campbell	COMPUTATIONAL TOOLS AND METHODS			2014	No
42	Google Academico	Trialstreamer: A living, automatically updated database of clinical trial reports	Randomized controlled trials (RCTs) are the gold standard method for evaluating whether a treatment works in health care but can be difficult to find and make use of. We describe the development and evaluation of a system to automatically find and categorize all new RCT reports	Iain J. Marshall, Benjamin Nye, Joel Kuiper, Anna Noel-Storr, Rachel Marshall, Rory Maclean, Frank Soboczenski, Ani Nenkova, James Thomas and Byron C. Wallace	Journal of the American Medical Informatics Association	King's College London, Northeastern University, University of Oxford, Vortext Systems, Cochrane Editorial and Methods Department, University of Pennsylvania, University College London	Estados Unidos, Reino Unido, Holanda	2020	Si

43	Google Academico	How to conduct systematic reviews more expeditiously?	Healthcare consumers, researchers, patients and policy makers increasingly use systematic reviews (SRs) to aid their decision-making process. However, the conduct of SRs can be a time-consuming and resource-intensive task. Often, clinical practice guideline developers or other decision-makers need to make informed decisions in a timely fashion (e.g. outbreaks of infection, hospital-based health technology assessments). Possible approaches to address the issue of timeliness in the production of SRs are to (a) implement process parallelisation, (b) adapt and apply innovative technologies, and/or (c) modify SR processes (e.g. study eligibility criteria, search sources, data extraction or quality assessment). Highly parallelised systematic reviewing requires substantial resources to support a team of experienced information specialists, reviewers and methodologists working alongside with clinical content experts to minimise the time for completing individual review steps while maximising the parallel progression of multiple steps. Effective coordination and management within the team and across external stakeholders are essential elements of this process. Emerging innovative technologies have a great potential for reducing workload and improving efficiency of SR production. The most promising areas of application would be to allow automation of specific SR tasks, in particular if these tasks are time consuming and resource intensive (e.g. language translation, study selection, data extraction). Modification of SR processes involves restricting, truncating and/or bypassing one or more SR steps, which may risk introducing bias to the review findings. Although the growing experiences in producing various types of rapid reviews (RR) and the accumulation of empirical studies exploring potential bias associated with specific SR tasks have contributed to the methodological development for expediting SR production, there is still a dearth of research examining the actual impact of methodological modifications and comparing the findings between RRs and SRs. This evidence would help to inform as to which SR tasks can be accelerated or truncated and to what degree, while maintaining the validity of review findings. Timely delivered SRs can be of value in informing healthcare decisions and recommendations, especially when there is practical urgency and there is no other relevant synthesised evidence.	Alexander Tsertsvadze, Yen-Fu Chen, David Moher, Paul Sutcliffe & Noel McCarthy	Systematic Reviews			2015	Si
44	Google Academico	Systematic review automation technologies	Systematic reviews, a cornerstone of evidence-based medicine, are not produced quickly enough to support clinical practice. The cost of production, availability of the requisite expertise and timeliness are often quoted as major contributors for the delay. This detailed survey of the state of the art of information systems designed to support or automate individual tasks in the systematic review, and in particular systematic reviews of randomized controlled clinical trials, reveals trends that see the convergence of several parallel research projects.	Guy Tsafnat, Paul Glasziou, Miew Keen Choong, Adam Dunn, Filippo	Systematic Reviews			2014	Disponible en Springer Link

			We surveyed literature describing informatics systems that support or automate the processes of systematic review or each of the tasks of the systematic review. Several projects focus on automating, simplifying and/or streamlining specific tasks of the systematic review. Some tasks are already fully automated while others are still largely manual. In this review, we describe each task and the effect that its automation would have on the entire systematic review process, summarize the existing information system support for each task, and highlight where further research is needed for realizing automation for the task. Integration of the systems that automate systematic review tasks may lead to a revised systematic review workflow. We envisage the optimized workflow will lead to system in which each systematic review is described as a computer program that automatically retrieves relevant trials, appraises them, extracts and synthesizes data, evaluates the risk of bias, performs meta-analysis calculations, and produces a report in real time.	Galgani & Enrico Coiera					
45	Google Academico	Embedding Randomized Controlled Trials into Medical Practice	To provide insight in how Pragmatic Clinical Trial (PCT) features can be embedded in RCTs and especially how registries and other routinely collected health data can be used to reduce costs, increase speed of recruitment and limit the amount of additional work needed for investigators, while keeping sufficient quality and acceptability.	Ouwens MJ	Value in Health		Suiza	2017	Disponible en Science Direct
46	Google Academico	Systematic review automation tools improve efficiency but lack of knowledge impedes their adoption: a survey	We investigated systematic review automation tool use by systematic reviewers, health technology assessors and clinical guideline developerst.	Anna Mae Scott, Connor Forbes, Justin Clark, Matt Carter, Paul Glasziou, Zachary Munn	Journal of Clinical Epidemiology	Bond University, The University of Adelaide	Australia	2021	Disponible en Science Direct
47	Google Academico	Tocilizumab and Systemic Corticosteroids in the Management of Patients with COVID-19: A Systematic Review and Meta-Analysis	Background: To date, there is no effective treatment for the new coronavirus disease (COVID-19). We aimed to systematically review the literature on the association between the combination of tocilizumab (TCZ) and systemic corticosteroid therapy (SCT) on outcomes of COVID-19 patients. Methods: We searched MEDLINE, Cochrane Central, and preprints, for studies in which health outcomes were compared between adults with severe COVID-19 who received TCZ and SCT and those who received standard of care without TCZ. Record screening, data extraction, and risk of bias assessment were performed in duplicate. Random effect models were used when pooling crude numbers and adjusted effect estimates of study outcomes. Results: Our search identified seventeen studies. The pooled crude mortality rate was lower in the combination arm (relative risk, RR=0.62, 95% confidence interval [CI]=0.42 – 0.91; I 2=60%). The adjusted mortality rates were also lower in the combination arm (RR=0.58, 95% CI=0.42 – 0.81; I	Hadeel Alkofide, Abdullah Almohaizeie, Sara Almuhaini, Bashayer Alotaibi, Khalid M.Alkharfy	International Journal of Infectious Diseases	King Saud University, King Saud University	Arabia Saudita	2021	Disponible en Science Direct

			2=71%). The rate of superinfections did not differ between the two interventions. Conclusions: The findings of this study show that combination of TCZ and SCT compared to SOC has lower mortality rates. There is an urgent need for well-designed randomized trials to assess the safety and efficacy of this combination in subjects with severe COVID-19.						
48	Google Academico	Semi-Automated evidence synthesis in health psychology: current methods and future prospects	The evidence base in health psychology is vast and growing rapidly. These factors make it difficult (and sometimes practically impossible) to consider all available evidence when making decisions about the state of knowledge on a given phenomenon (e.g., associations of variables, effects of interventions on particular outcomes). Systematic reviews, meta-analyses, and other rigorous syntheses of the research mitigate this problem by providing concise, actionable summaries of knowledge in a given area of study. Yet, conducting these syntheses has grown increasingly laborious owing to the fast accumulation of new evidence; existing, manual methods for synthesis do not scale well. In this article, we discuss how semi-automation via machine learning and natural language processing methods may help researchers and practitioners to review evidence more efficiently. We outline concrete examples in health psychology, highlighting practical, open-source technologies available now. We indicate the potential of more advanced methods and discuss how to avoid the pitfalls of automated reviews.	Iain J. Marshall,Blair T. Johnson,Zigeng Wang,Sanguthevar Rajasekaran &Byron C. Wallace	Health Psychology Review			2020	Si
49	Google Academico	Modernizing the systematic review process to inform comparative effectiveness: tools and methods	Systematic reviews are being increasingly used to inform all levels of healthcare, from bedside decisions to policy-making. Since they are designed to minimize bias and subjectivity, they are a preferred option to assess the comparative effectiveness and safety of healthcare interventions. However, producing systematic reviews and keeping them up-to-date is becoming increasingly onerous for three reasons. First, the body of biomedical literature is expanding exponentially with no indication of slowing down. Second, as systematic reviews gain wide acceptance, they are also being used to address more complex questions (e.g., evaluating the comparative effectiveness of many interventions together rather than focusing only on pairs of interventions). Third, the standards for performing systematic reviews have become substantially more rigorous over time. To address these challenges, we must carefully prioritize the questions that should be addressed by systematic reviews and optimize the processes of research synthesis. In addition to reducing the workload involved in planning and conducting systematic reviews, we also need to make efforts to increase the transparency, reliability and validity of the review process; these aims can be grouped under the umbrella of 'modernization' of the systematic review process.	Byron C Wallace, Issa J Dahabreh, Christopher H Schmid, Joseph Lau & Thomas A Trikalinos	JOURNAL OF COMPARATIVE EFFECTIVENESS RESEARCH			2013	Si

50	Google Academico	Machine Learning in Evidence Synthesis Research	In this chapter, we will explore how Systemic Reviews (SR) are traditionally conducted and how the process of arriving at a valuable SR can be made more efficient and less error prone using Machine Learning (ML) techniques. As the integration of ML at the screening stage of SRs has reached the highest level of maturity, we will explain the techniques utilized. We will further describe the extraction process from primary studies supported by ML techniques. The discussion of pitfalls when conducting SRs concludes the chapter, specifically how ML can address bias. Lastly, we address the inherent limitations of artificial intelligence in healthcare with a special emphasis on ML for the use in SRs.	Alonso Carrasco-Labra, Olivia Urquhart, Heiko Spallek	Machine Learning in Dentistry			2021	Si
51	Google Academico	Combined Pharmacotherapy and Cognitive Behavioral Therapy for Adults With Alcohol or Substance Use Disorders A Systematic Review and Meta-analysis	Substance use disorders (SUDs) represent a pressing public health concern. Combined behavioral and pharmacological interventions are considered best practices for addiction. Cognitive behavioral therapy (CBT) is a first-line intervention, yet the superiority of CBT compared with other behavioral treatments when combined with pharmacotherapy remains unclear. An understanding of the effects of combined CBT and pharmacotherapy will inform best-practice guidelines for treatment of SUD.	Lara A. Ray; Lindsay R. Meredith; Brian D. Kiluk; Justin Walthers; Kathleen M. Carroll; Molly Magill	JAMA Netw Open			2020	Si
52	Google Academico	Identifying Attributes of Cancer Treatments: What Do Stakeholders Consider Important	Increasing cancer therapy costs have produced a recent emergence of numerous value frameworks in oncology. However, variation within frameworks and lack of clarity regarding inclusion criteria for value 'attributes' presents a need for further research. This study aims to identify such attributes based on the priorities of different stakeholder groups in order to develop a new conceptual framework for value in oncology	Baba C, Briggs A, Mcintosh E, Lipitz-Snyderman A, Kaltenboeck A, Bach PB	VALUE IN HEALTH	University of Glasgow, Memorial Sloan Kettering Cancer Center	Reino Unido, Estados Unidos	2017	Si
53	Google Academico	Tools to support the automation of systematic reviews: a scoping review	The objectives of this scoping review are to identify the reliability and validity of the available tools, their limitations and any recommendations to further improve the use of these tools.	Hanan Khalil, Daniel Ameen, Amrita Zarnegara	Journal of Clinical Epidemiology	La Trobe University, Monash University, Swinburne University	Australia	2022	Disponible en Science Direct
54	Google Academico	Effects of Structured Exercise Interventions for Older Adults Hospitalized with Acute Medical Illness: A Systematic Review	This review examined effects of structured exercise (aerobic walking, with or without complementary modes of exercise) on cardiorespiratory measures, mobility, functional status, healthcare utilization, and Quality of Life in older adults (≥60 years) hospitalized for acute medical illness. Inclusion required exercise protocol, at least one patient-level or utilization outcome, and at least one physical assessment point during hospitalization or within 1 month of intervention. MEDLINE, Embase, and CINAHL databases were searched for studies published from 2000 to March 2015. Qualitative synthesis of 12 articles, reporting on 11 randomized controlled (RCT)	Frances A. Kanach, Amy M. Pastva, Katherine S. Hall, Juliessa M. Pavon, Miriam C. Morey	J Aging Phys	Duke University	Estados Unidos	2018	Si

			and quasi-experimental studies described a heterogeneous set of exercise programs and reported mixed results across outcome categories. Methodological quality was independently assessed by 2 reviewers using the Cochrane Collaboration Risk of Bias tool. Larger, well-designed RCTs are needed, incorporating measurement of pre-morbid function, randomization with intention-to-treat analysis, examination of a targeted intervention with pre-defined intensity, and reported adherence and attrition.						
55	Google Academico	Anthropometric Outcomes of Children and Adolescents Using Telehealth with Weight Management Interventions Compared to Usual Care: A Systematic Review and Meta-analysis	This systematic review and meta-analysis evaluated the effect of telehealth (TH) weight management interventions compared to usual care on anthropometric outcomes in children and adolescents with overweight and obesity.	Cheryl A Margetin, Diane Rigassio Radler, Kyle Thompson, Jane Ziegler, Margaret Dreker, Laura Byham-Gray & Mei Chung	JOURNAL OF THE AMERICAN COLLEGE OF NUTRITION	Rutgers University Biomedical and Health Sciences, Seton Hall University, Appalachian State University, Tufts University School of Medicine	Estados Unidos	2020	Si
56	Google Academico	Motivational Interviewing to Treat Adolescents With Obesity: A Meta-analysis	Conduct a systematic review of MI for treating overweight and obesity in adolescents and meta-analysis of its effects on anthropometric and cardiometabolic outcomes.	Monique K. Vallabhan, Elizabeth Y. Jimenez, Jacob L. Nash, Diana Gonzales-Pacheco, Kathryn E. Coakley, Shelly R. Noe, Conni J. DeBlieck, Linda C. Summers, Sarah W. Feldstein-Ewing, Alberta S. Kong	Pediatrics	University of New Mexico	Estados Unidos	2018	Si

57	Google Academico	Comparative Effectiveness of Combining MTX with Biologic Drug Therapy Versus Either MTX or Biologics Alone for Early Rheumatoid Arthritis in Adults: a Systematic Review and Network Meta-analysis	Comparative effectiveness of early rheumatoid arthritis (RA) treatments remains uncertain.	Katrina E. Donahue M.D., M.P.H., Elizabeth R. Schulman M.D., Gerald Gartlehner M.D., M.P.H., Beth L. Jonas M.D., F.A.C.R., Emmanuel Coker-Schwimmer M.P.H., Sheila V. Patel B.S.P.H., Rachel Palmieri Weber, Carla M. Bann, Meera Viswanathan	Journal of General Internal Medicine			2019	Disponible en Springer Link
58	Google Academico	Prospective approaches to accumulating evidence	This chapter covers methods for keeping abreast of the accumulating evidence to help a review team understand when a systematic review might need updating. This includes the processes that can be put into place to monitor relevant publications, and algorithms that have been proposed to determine whether or when it is appropriate to revisit the review to incorporate new findings. The chapter outlines a vision for regularly updated reviews, known as 'living' systematic reviews, which are continually updated, with new evidence being identified and incorporated as soon as it becomes available. It also addresses concerns about the regular repeating of statistical tests in meta-analyses as they are updated over time. Cochrane actively discourages use of the notion of statistical significance in favour of reporting estimates and confidence intervals. Nevertheless, sequential approaches are an established method in randomized trials, and may play a role in a prospectively planned series of trials in a prospective meta-analysis.	James Thomas, Lisa M Askie, Jesse A Berlin, Julian H Elliott, Davina Ghersi, Mark Simmonds, Yemisi Takwoingi, Jayne F Tierney, Julian PT	Systematic Reviews of Interventions			2019	Si

				Higgins					
59	Google Academico	Using text mining for study identification in systematic reviews: a systematic review of current approaches	The large and growing number of published studies, and their increasing rate of publication, makes the task of identifying relevant studies in an unbiased way for inclusion in systematic reviews both complex and time consuming. Text mining has been offered as a potential solution: through automating some of the screening process, reviewer time can be saved. The evidence base around the use of text mining for screening has not yet been pulled together systematically; this systematic review fills that research gap. Focusing mainly on non-technical issues, the review aims to increase awareness of the potential of these technologies and promote further collaborative research between the computer science and systematic review communities.	Alison O'Mara-Eves, James Thomas, John McNaught, Makoto Miwa & Sophia Ananiadou	Systematic Reviews			2015	Si
60	Google Academico	Improving reference prioritisation with PICO recognition	Machine learning can assist with multiple tasks during systematic reviews to facilitate the rapid retrieval of relevant references during screening and to identify and extract information relevant to the study characteristics, which include the PICO elements of patient/population, intervention, comparator, and outcomes. The latter requires techniques for identifying and categorising fragments of text, known as named entity recognition.	Austin J. Brockmeier, Meizhi Ju, Piotr Przybyła & Sophia Ananiadou	BMC Medical Informatics and Decision Making			2019	Si
61	Google Academico	A meta-analysis of cognitive-behavioral therapy for alcohol or other drug use disorders: Treatment efficacy by contrast condition	Objective: This meta-analysis examined 30 randomized controlled trials (32 study sites; 35 study arms) that tested the efficacy of cognitive-behavioral therapy (CBT) for alcohol or other drug use disorders. The study aim was to provide estimates of efficacy against three levels of experimental contrast (i.e., minimal [k = 5]; nonspecific therapy [k = 11]; specific therapy [k = 19]) for consumption frequency and quantity outcomes at early (1 to 6 months [kes = 41]) and late (8+ months [kes = 26]) follow-up time points. When pooled effect sizes were statistically heterogeneous, study-level moderators were examined. Method: The inverse-variance weighted effect size was calculated for each study and pooled under random effects assumptions. Sensitivity analyses included tests of heterogeneity, study influence, and publication bias. Results: CBT in contrast to minimal treatment showed a moderate and significant effect size that was consistent across outcome type and follow-up. When CBT was contrasted with a nonspecific therapy or treatment as usual, treatment effect was statistically significant for consumption frequency and quantity at early, but not late, follow-up. CBT effects in contrast to a specific therapy were consistently nonsignificant across outcomes and follow-up time points. Of 10 pooled effect sizes examined, two showed moderate heterogeneity, but multivariate analyses revealed few systematic predictors of between-study variance. Conclusions: The current meta-analysis	Magill, Molly Ray, Lara Kiluk, Brian Hoadley, Ariel Bernstein, Michael Tonigan, J. Scott Carroll, Kathleen	Journal of Consulting and Clinical Psychology			2019	Si

			shows that CBT is more effective than a no treatment, minimal treatment, or nonspecific control. Consistent with findings on other evidence-based therapies, CBT did not show superior efficacy in contrast to another specific modality. (PsycInfo Database Record (c) 2020 APA, all rights reserved)						
62	Google Academico	Associations Between Tooth Loss, With or Without Dental Prostheses, and Malnutrition Risk in Older Adults	A systematic review was conducted to investigate the associations between missing teeth, teeth replaced with dental prostheses, and malnutrition risk, as determined by the Mini Nutritional Assessment (MNA), among community-dwelling older adults. A comprehensive search strategy was designed to identify studies published between 2000 and 2014 that met inclusion criteria. Searches were conducted in Medline, CINAHL, and Cochrane Libraries. Preferred Reporting Items for Systematic Reviews and Meta Analyses guidelines were followed and risk-of-bias analyses were performed. Study data were abstracted and synthesized in narrative and summary tables. Of the 8 studies that met inclusion criteria, 5 found significant associations between MNA score and tooth loss. Those with fewer teeth/limited occlusion had lower MNA scores, hence were at greater risk for malnutrition than those with a greater number of teeth and/or more occluding teeth pairs. Provision of dentures to older adults was associated with significant increases in MNA scores. The studies reviewed support associations between tooth loss, replacement with prostheses, and malnutrition risk; individuals with more missing teeth and limited occlusion were at increased risk of malnutrition. Additional high-quality research seems warranted to substantiate a causal relationship between dental and nutritional status. Inadequate oral health and poor nutritional status are associated with increased mortality risk in older adults. Efforts to optimize oral health and nutritional status of older adults are needed.	Zelig, Rena; Touger-Decker, Riva; Chung, Mei; Byham-Gray, Laura	Topics in Clinical Nutrition			2019	Si
63	Google Academico	Cognitive-Behavioral Interventions Targeting Alcohol or Other Drug Use and Co-Occurring Mental Health Disorders: A Meta-Analysis	This meta-analysis reviewed 15 clinical trials (18 study sites/arms), examining the efficacy of an integrated cognitive-behavioral intervention (CBI) delivered to individuals with an alcohol or other drug use disorder and a co-occurring mental health disorder (AOD/MHD). Outcomes were alcohol or other drug use and mental health symptoms at post-treatment through follow-up.	Kahini Mehta, Ariel Hoadley, Lara A Ray, Brian D Kiluk, Kathleen M Carroll, Molly Magill	Alcohol and Alcoholism			2021	Si

64	Google Academico	The Effectiveness of Regulations and Behavioral Interventions on Head Impacts and Concussions in Youth, High-School, and Collegiate Football: A Systematized Review	The purpose of this study was to assess the effectiveness of regulations and behavioral interventions on head impacts and concussions in youth, high-school, and collegiate football, using a systematic search strategy to identify relevant literature. Six databases were searched using key search terms related to three categories: football, head-injuries, and interventions. Studies that met inclusion criteria were included in the study and underwent data extraction. Twenty articles met inclusion criteria and were included in the final systematized review. Of the 20 included studies, 8 studies evaluated interventions in high-school football, 5 studies evaluated interventions in collegiate football, 6 studies evaluated interventions in youth football, and 1 study evaluated interventions in both, high-school and collegiate football. The four categories of interventions and regulations included rule changes, training, education/instruction/coaching tactical changes, and tackle football alternatives. Studies evaluating the effectiveness of interventions and regulations on reducing head impact exposures or head injuries have shown mixed results. Some regulations may be more effective than others, but methodological design and risk of bias pose limitations to generalize effects.	Nicole Phillips & Joseph J. Crisco	Annals of Biomedical Engineering			2020	Si
65	Google Academico	Heterogeneity of outcomes for intraoperative music interventions: a scoping review and evidence map	Over the past 30 years, numerous studies have been performed that assess the efficacy of intraoperative music as an adjunctive means to regional and local anaesthesia to improve clinical outcomes. Despite an emerging body of evidence and growing adoption of music in surgical settings, the variety of interventions studied, and the heterogeneity of outcomes and outcome measurement tools applied makes difficult the task of aggregating evidence.	Melanie Ambler, Stacey Springs, Dioscaris Garcia, Christopher Born	Evidence synthesis	Brown University Division of Biology and Medicine, Diane Weiss Center for Orthopaedic Trauma Research, Center for Evidence Synthesis, Brown University School of Public Health,	Estados Unidos		No
66	Google Academico	Efficacy and Safety Profile of Anti–tumor Necrosis Factor-α Versus Anti-integrin Agents for the Treatment of Crohn's Disease: A Network Meta-analysis of Indirect Comparisons	To compare the benefits and harms of anti–tumor necrosis factor (TNF)-α and anti-integrin agents as induction and maintenance therapy in adult patients with Crohn's disease.	Michael Miligkos, Konstantinos Papamichael, NielsVande CasteeleP harm, Gerassimos J. Mantzaris, Ann Gils,	Clinical Therapeutics	University of Thessaly School of Medicine, Evaggelismos Hospital, KU Leuven, University of California, Tufts University School of Medicine	Grecia, Belgica, Estados Unidos	2016	Si

				Barrett G. Levesque, Elias Zintzaras					
67	Google Academico	Small Improvements in Working Memory After Cognitive Training do not Transfer to Fluid Intelligence: Evidence From a Meta-analysis	Meta-analyses have found that working memory (WM) can be improved with WM training programmes, however some authors have suggested that these improvements are mostly driven by biases in the measurement of WM, more specifically, by using similar tasks for assessment and training. In the present meta-analysis, we investigate if WM can be improved in	Jose A. RodasCiara Greene					Si
68	Google Academico	Evidence Brief: Suicide Prevention in Veterans	Despite the US Department of Veterans Affairs' (VA) increased efforts over the past decade in implementing comprehensive Suicide Prevention Program initiatives, according to the new VA National Suicide Data Report 2005-2015, an average of 20 Veterans continue to die each day by suicide. An important barrier to the success of VA's suicide prevention initiatives may be the lack of adequate evidence in Veterans supporting recommendations of any specific risk assessment method or prevention intervention.	Peterson K, Anderson J, Bourne D				2019	No
69	Google Academico	Evidence Brief: Suicide Prevention in Veterans		Kim Peterson, Johanna Anderson, Donald Bourne				2018	Si
70	Science direct	A systematic review of machine learning and automation in burn wound evaluation: A promising but developing frontier	Visual evaluation is the most common method of evaluating burn wounds. Its subjective nature can lead to inaccurate diagnoses and inappropriate burn center referrals. Machine learning may provide an objective solution. The objective of this study is to summarize the literature on ML in burn wound evaluation.	Samantha Huanga, Justin Dang,Clifford C. Sheckter, Haig A. Yenikomshian, Justin Gillenwater	Burns	University of Southern California, University of Southern California, Stanford University	Estados Unidos	2021	Si
71	Science direct	Technology-assisted risk of bias assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool	In this prospective cross-sectional evaluation, we used RobotReviewer to assess risk of bias among 1,180 trials. We computed reliability with human reviewers using Cohen's kappa coefficient and calculated sensitivity and specificity. We investigated differences in reliability by risk of bias domain, topic, and outcome type using the chi-square test in meta-analysis.	Allison Gates, Ben Vandermeer, Lisa Hartling	Journal of Clinical Epidemiology	University of Alberta	Canada	2018	Si
72	Science direct	Tools to support the automation of systematic reviews: a	The objectives of this scoping review are to identify the reliability and validity of the available tools, their limitations and any recommendations to further improve the use of these tools.	Hanan Khalila, Daniel	Journal of Clinical Epidemiology	La Trobe University, Monash	Australia	2022	Si

		scoping review		Ameen, Amrita Zarnegara		University, Swinburne University			
73	Scienc e direct	Machine learning applications to neuroimaging for glioma detection and classification: An artificial intelligence augmented systematic review	Glioma is the most common primary intraparenchymal tumor of the brain and the 5-year survival rate of high-grade glioma is poor. Magnetic resonance imaging (MRI) is essential for detecting, characterizing and monitoring brain tumors but definitive diagnosis still relies on surgical pathology. Machine learning has been applied to the analysis of MRI data in glioma research and has the potential to change clinical practice and improve patient outcomes. This systematic review synthesizes and analyzes the current state of machine learning applications to glioma MRI data and explores the use of machine learning for systematic review automation. Various datapoints were extracted from the 153 studies that met inclusion criteria and analyzed. Natural language processing (NLP) analysis involved keyword extraction, topic modeling and document classification. Machine learning has been applied to tumor grading and diagnosis, tumor segmentation, non-invasive genomic biomarker identification, detection of progression and patient survival prediction. Model performance was generally strong (AUC = 0.87 ± 0.09; sensitivity = 0.87 ± 0.10; specificity = 0.0.86 ± 0.10; precision = 0.88 ± 0.11). Convolutional neural network, support vector machine and random forest algorithms were top performers. Deep learning document classifiers yielded acceptable performance (mean 5-fold cross-validation AUC = 0.71). Machine learning tools and data resources were synthesized and summarized to facilitate future research. Machine learning has been widely applied to the processing of MRI data in glioma research and has demonstrated substantial utility. NLP and transfer learning resources enabled the successful development of a replicable method for automating the systematic review article screening process, which has potential for shortening the time from discovery to clinical application in medicine.	Quinlan D. Buchlaka, Nazanin Esmailia, Jean-Christophe Leveque,C hristine Bennetta, Farrokh Farrokhi, Massimo Piccardi	Journal of Clinical Neuroscience	The University of Notre Dame Australia, University of Technology Sydney, Virginia Mason Medical Center	Estados Unidos	20 21	Si
74	Scienc e direct	Systematic review automation tools improve efficiency but lack of knowledge impedes their adoption: a survey	We investigated systematic review automation tool use by systematic reviewers, health technology assessors and clinical guideline developerst.	Anna Mae Scotta, Connor Forbesa, Justin Clarka, Matt Cartera, Paul Glaszioua, Zachary Munn	Journal of Clinical Epidemiology	Bond University, The University of Adelaide	Australia	20 21	Si

75	Science direct	Living systematic reviews: 2. Combining human and machine effort	New approaches to evidence synthesis, which use human effort and machine automation in mutually reinforcing ways, can enhance the feasibility and sustainability of living systematic reviews. Human effort is a scarce and valuable resource, required when automation is impossible or undesirable, and includes contributions from online communities ("crowds") as well as more conventional contributions from review authors and information specialists. Automation can assist with some systematic review tasks, including searching, eligibility assessment, identification and retrieval of full-text reports, extraction of data, and risk of bias assessment. Workflows can be developed in which human effort and machine automation can each enable the other to operate in more effective and efficient ways, offering substantial enhancement to the productivity of systematic reviews. This paper describes and discusses the potential—and limitations—of new ways of undertaking specific tasks in living systematic reviews, identifying areas where these human/machine "technologies" are already in use, and where further research and development is needed. While the context is living systematic reviews, many of these enabling technologies apply equally to standard approaches to systematic reviewing.	James Thomas, Anna Noel-Storrb, Iain Marshall, Byron Wallace, Steven McDonald, Chris Mavergames, Paul Glasziou, Ian Shemilt, Anneliese Synnote, Tari Turner,Julian Elliott	Journal of Clinical Epidemiology	University College London, University of Oxford, Kings College, Northeastern University, Monash University, Bond University, La Trobe University, Monash University and Alfred Hospital	Estados Unidos	2017	Si
76	Science direct	Recommendations for assessing the risk of bias in systematic reviews of health-care interventions	Risk-of-bias assessment is a central component of systematic reviews, but little conclusive empirical evidence exists on the validity of such assessments. In the context of such uncertainty, we present pragmatic recommendations that promote transparency and reproducibility in processes, address methodological advances in the risk-of-bias assessment, and can be applied consistently across review topics.	Meera Viswanathana, Carrie D. Patnode, Nancy D. Berkman, Eric B. Bass,Stephanie Chang,Lisa Hartlinge, M. Hassan Murad, Jonathan R. Treadwell, Robert L. Kaneh	Journal of Clinical Epidemiology		Estados Unidos	2018	Si
77	Science direct	Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a	This study developed, calibrated, and evaluated a machine learning classifier designed to reduce study identification workload in Cochrane for producing systematic reviews.	James Thomasa, Steve McDonald, Anna Noel-Storr	Journal of Clinical Epidemiology	University College London, Monash University, University of Oxford, Kings	Australia, Reino Unido	2021	Si

		randomized controlled trial classifier for Cochrane Reviews		,Ian Shemilt,Julian Elliott, Chris Mavergames, Iain J. Marshall		College London			
78	Science direct	Diagnostic and prognostic prediction models in ventilator-associated pneumonia: Systematic review and meta-analysis of prediction modelling studies	Existing expert systems have not improved the diagnostic accuracy of ventilator-associated pneumonia (VAP). The aim of this systematic literature review was to review and summarize state-of-the-art prediction models detecting or predicting VAP from exhaled breath, patient reports and demographic and clinical characteristics.	Tuomas Frondelius, Irina Atkova,Jouko Miettunence, Jordi Rellod,Miia M. Jansson	Journal of Critical Care		Finlandia, España, Francia	2022	Si
79	Science direct	Perspectives: A surgeon's guide to machine learning	The exponential increase in the volume and complexity of healthcare data presents new challenges to researchers and clinicians in analysis and interpretation. The requirement for new strategies to extract meaningful information from large, noisy datasets has led to the development of the field of big data analytics. Artificial intelligence (AI) is a general-purpose technology in which machines carry out tasks traditionally thought to be only achievable by humans. Machine learning (ML) is an approach to AI in which machines can “learn” to perform tasks in an automated process, rather than being explicitly programmed by a human. Research aiming to apply ML techniques to classification, prediction and decision-making problems in healthcare has increased 61-fold from 2005 to 2019, mirroring this sense of early promise. The field of healthcare ML is relatively young, and many critical steps are needed before adoption into clinical practice, including transparent, unbiased development and reporting of algorithms. Articles claiming that machines can outperform, or replace, doctors in high-level tasks, such as diagnosis or prognostication, must be carefully appraised. It is critical that surgeons have an understanding of the principles and terminology of AI and ML to evaluate these claims and to take an active role in directing research. This article is an up-to-date review and primer for surgeons covering the core tenets of ML applied to surgical problems, including algorithm types and selection, model training and validation, interpretation of common outcome metrics, current and future reporting guidelines and discussion of the challenges and limitations in this field.	Rachel Y. L. Kuo, Conrad J. Harrison, Benjamin E. Jones, Luke Geoghegan, Dominic Furniss	International Journal of Surgery	University of Oxford, Imperial College London	Reino Unido	2021	Si
80	Science direct	A systematic review on artificial intelligence in robot-assisted surgery	Despite the extensive published literature on the significant potential of artificial intelligence (AI) there are no reports on its efficacy in improving patient safety in robot-assisted surgery (RAS). The purposes of this work are to systematically review the published literature on AI in RAS, and to identify and discuss current limitations	Andrea Moglia, Konstantinos Georgiou,	International Journal of Surgery	University of Pisa, Kapodistrian University of Athens,	Italia, Reino Unido, Grecia, Estados Unidos	2021	Si

			and challenges.	Evangelos Georgiou, Richard M. Satavad, Alfred Cuschieri		University of Washington Medical Center, University of Dundee			
81	Science direct	Diagnostic utility of artificial intelligence for left ventricular scar identification using cardiac magnetic resonance imaging—A systematic review	Accurate, rapid quantification of ventricular scar using cardiac magnetic resonance imaging (CMR) carries importance in arrhythmia management and patient prognosis. Artificial intelligence (AI) has been applied to other radiological challenges with success	NikeshJathanna, AnnaPodlasek, Albert Sokol, DorotheeAuer, XinChen, ShahnazJamil-Copley	Cardiovascular Digital Health Journal	University of Nottingham	Reino Unido	2021	Si
82	Science direct	Artificial intelligence models for tooth-supported fixed and removable prosthodontics: A systematic review	Artificial intelligence applications are increasing in prosthodontics. Still, the current development and performance of artificial intelligence in prosthodontic applications has not yet been systematically documented and analyzed.	Marta Revilla-León, Miguel Gómez-Polo, Shantanu Vyasc,Abdul Basir Barmak, German O. Gallucci, Wael Att, Mutlu Özcan, Vinayak R. Krishnamurthy	The Journal of Prosthetic Dentistry	Texas A&M University, Complutense University of Madrid, University of Rochester Medical Center, Harvard School of Dental Medicine, University of Zürich	Estados Unidos, Suiza, España	2021	Si
83	Science direct	Enhancing WIKISTIM.org Using Machine Learning Approaches		Aaron Lawson McLean, Jan Walter	Neuromodulation: Technology at the Neural Interface	Jena University Hospital	Alemania	2019	No
84	Science direct	The PRISMA 2020 statement: An updated guideline for reporting systematic reviews	The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement, published in 2009, was designed to help systematic reviewers transparently report why the review was done, what the authors did, and what they found. Over the past decade, advances in systematic review methodology and terminology have necessitated an update to the guideline. The	Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt,	Journal of Clinical Epidemiology	Monash University, University of Amsterdam, Université de Paris, Bond	Australia, Francia, Estados, Canada, Libano, Holanda	2021	Si

			PRISMA 2020 statement replaces the 2009 statement and includes new reporting guidance that reflects advances in methods to identify, select, appraise, and synthesise studies. The structure and presentation of the items have been modified to facilitate implementation. In this article, we present the PRISMA 2020 27-item checklist, an expanded checklist that details reporting recommendations for each item, the PRISMA 2020 abstract checklist, and the revised flow diagrams for original and updated reviews.	Isabelle Boutron, Tammy C. Hoffmann, Cynthia D.Mulrowe , Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akli, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalurst, Tianjing Liuv, Elizabeth W. Loderwx , Evan Mayo-Wilsony, David Moher		University, University of Texas, University of Ottawa, American University of Beirut, McMaster University, University of York			
85	Science direct	Developments, application, and performance of artificial intelligence in dentistry – A systematic review	Artificial intelligence (AI) has made deep inroads into dentistry in the last few years. The aim of this systematic review was to identify the development of AI applications that are widely employed in dentistry and evaluate their performance in terms of diagnosis, clinical decision-making, and predicting the prognosis of the treatment.	Sanjeev B. Khanagar, AliA l-ehaideb, Prabhadevi C.Maganur, Satish Vishwanathaiah, Shankargouda Patile, Hosam A.	Journal of Dental Sciences	King Saud Bin Abdulaziz University, Jazan University,	Arabia Saudita	2021	Si

				Baeshen, Sachin C. Sarode, Shilpa Bhandi					
86	Scienc e direct	The PRISMA 2020 statement: An updated guideline for reporting systematic reviews	The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement, published in 2009, was designed to help systematic reviewers transparently report why the review was done, what the authors did, and what they found. Over the past decade, advances in systematic review methodology and terminology have necessitated an update to the guideline. The PRISMA 2020 statement replaces the 2009 statement and includes new reporting guidance that reflects advances in methods to identify, select, appraise, and synthesise studies. The structure and presentation of the items have been modified to facilitate implementation. In this article, we present the PRISMA 2020 27-item checklist, an expanded checklist that details reporting recommendations for each item, the PRISMA 2020 abstract checklist, and the revised flow diagrams for original and updated reviews.	Matthew J. Pagea, Joanne E. McKenzie, Patrick M.Bossuyt , Isabelle Boutron, Tammy C. Hoffmann	International Journal of Surgery	Monash Universit, Amsterdam University Medical Centres, Université de Paris, Bond University	Australia, Holanda, Francia	20 21	Si
87	Scienc e direct	Diagnostic Test Accuracy of Deep Learning Detection of COVID-19: A Systematic Review and Meta-Analysis	To perform a meta-analysis to compare the diagnostic test accuracy (DTA) of deep learning (DL) in detecting coronavirus disease 2019 (COVID-19), and to investigate how network architecture and type of datasets affect DL performance.	Temitope Emmanuel Komolafe, YuzhuCao , Benedictor Alexander Nguchu, PatriceMo nkam, Ebenezer Obaloluwa Olaniyi, HaotianSu n, JianZheng , Xiaodong Yang	Academic Radiology		China	20 21	Disponib le en Clinical Key
88	Scienc e direct	Assessing risk of bias in studies that evaluate health care interventions: recommendations in the misinformation age	Methods to assess the risk of bias in a way that is reliable, reproducible. and transparent to readers, have evolved over time. Viswanathan et al. recently provided updated recommendations for assessing risk of bias in systematic reviews of health care interventions. We comment on their recommendations and discuss new tools in development that we, as co-convenors and coordinators of the Cochrane Bias Methods Group, are leading, which complement the methods recommended.	Matthew J. Pagea, Isabelle Boutron, Camilla Hansen, Douglas G.	Journal of Clinical Epidemiology	Monash University, Paris Descartes University, University of Southern Denmark & Odense	Reino Unido, Dinamarca, Francia, Australia	20 18	Si

				Altmang, Asbjørn Hróbjartsson		University Hospital, University of Oxford			
89	Science direct	PRM88 - THE ADVANCEMENT OF TOOLS FOR AUTOMATING DATA EXTRACTION IN SYSTEMATIC REVIEWS	The potential for automation to reduce time required to undertake the key stages of systematic reviews has seen a growth in software applications in recent years. A recent systematic review found no unified information extraction framework for automating data extraction in systematic reviews. Whilst a proliferation of risk of bias and "process management" tools exist, data extraction from text and tables in systematic review has received less attention. We reviewed available tools and evaluative studies to determine progress in automation in this area.	D. A. Scott, J. Colquitt, E. Loveman, P. Royle	Value in Health	University of Leicester, Warwick University	Reino Unido	2018	Si
90	Science direct	Three-dimensional cone beam computed tomography analysis protocols for condylar remodelling following orthognathic surgery: a systematic review	Orthognathic surgery involving the mandible can lead to remodelling of the temporomandibular joint (TMJ). Cone beam computed tomography (CBCT) provides an easily accessible three-dimensional (3D) approach to study this entity. A systematic review of the literature was performed with the aim of identifying condylar remodelling analysis protocols using CBCT-derived 3D models. The search yielded 10 eligible studies. The systematic review identified three pillars of a condylar remodelling analysis protocol that were retrievable from each of the included studies: (1) registration, (2) segmentation, and (3) analysis. The studies lacked consensus on how these pillars should be transferred to their respective protocol. Through critical assessment, criteria for a universal condylar remodelling analysis are suggested: (1) performance of a regional voxel-based registration of baseline and postoperative CBCT scans using an anatomical region not prone to postoperative changes, (2) application of a (semi-)automated 3D segmentation algorithm, (3) performance of a combination of both volumetric and surface-based 3D condylar analysis, and (4) extensive validation of each step of the protocol. The homogenization of condylar remodelling analysis protocols and their incorporation into virtual planning software suites raises the potential for the inclusion of larger numbers of patients in future prospective studies in order to gain evidence-based data.	P. J. Verhelst, L. Verstraete, E. Shaheen, S. Shujaat, V. Darche, R. Jacobs, G. Swennen, C. Politis	International Journal of Oral and Maxillofacial Surgery	University Hospitals Leuven,	Reino Unido, Belgica, Suecia	2020	Si
91	Science direct	Evaluation of a semi-automated data extraction tool for public health literature-based reviews: Dextr	There has been limited development and uptake of machine-learning methods to automate data extraction for literature-based assessments. Although advanced extraction approaches have been applied to some clinical research reviews, existing methods are not well suited for addressing toxicology or environmental health questions due to unique data needs to support reviews in these fields.	Vickie R. Walker, Charles P. Schmitt, Mary S. Wolfe, Artur J. Nowak, Kuba Kulesza, Ashley R. Williams, Rob Shin,	Environment International		Estados Unidos	2022	Si

				Jonathan Cohen, Dave Burch, Matthew D. Stout, Kelly A. Shipkowski, Andrew A. Rooney					
92	Science direct	Osseointegration Pharmacology: A Systematic Mapping Using Artificial Intelligence	Clinical performance of osseointegrated implants could be compromised by the medications taken by patients. The effect of a specific medication on osseointegration can be easily investigated using traditional systematic reviews. However, assessment of all known medications requires the use of evidence mapping methods. These methods allow assessment of complex questions, but they are very resource intensive when done manually. The objective of this study was to develop a machine learning algorithm to automatically map the literature assessing the effect of medications on osseointegration. Datasets of articles classified manually were used to train a machine-learning algorithm based on Support Vector Machines. The algorithm was then validated and used to screen 599,604 articles identified with an extremely sensitive search strategy. The algorithm included 281 relevant articles that described the effect of 31 different drugs on osseointegration. This approach achieved an accuracy of 95%, and compared to manual screening, it reduced the workload by 93%. The systematic mapping revealed that the treatment outcomes of osseointegrated medical devices could be influenced by drugs affecting homeostasis, inflammation, cell proliferation and bone remodeling. The effect of all known medications on the performance of osseointegrated medical devices can be assessed using evidence mappings executed with highly accurate machine learning algorithms.	Mohammed Mahria, Nicole Shena, Francisco Berrizbeitia, Rania Rodana, Ammar Daer, Matthew Faigana, Doaa Taqia, Kevin Yang Wu, Motahareh Ahmadia, Maxime Ducreta, Elham Emamia, Faleh Tamimi	Acta Biomaterialia	McGill University, Jazan University, Concordia University, Laval University, Qatar University	Canada, Arabia Saudita, Jordan, Francia, Qatar	2021	Si
93	Science direct	SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation	In the screening phase of systematic review, researchers use detailed inclusion/exclusion criteria to decide whether each article in a set of candidate articles is relevant to the research question under consideration. A typical review may require screening thousands or tens of thousands of articles in and can utilize hundreds of person-hours of labor.	Brian E. Howarda, Jason Phillipsa, Arpit Tandona, Adyasha Maharanaa, Rebecca Elmorea, Deepak Mava, Alex	Environment International		Estados Unidos	2020	Si

				Sedykha, Kristina Thayer, B. Alex Merrick, Vickie Walker, Andrew Rooney, Ruchir R. Shah					
94	Science direct	Development of machine translation technology for assisting health communication: A systematic review	To (1) characterize how machine translation (MT) is being developed to overcome language barriers in health settings; and (2) based on evaluations presented in the literature, determine which MT approaches show evidence of promise and what steps need to be taken to encourage adoption of MT technologies in health settings.	Kristin N. Dewa, Anne M. Turnerb, Yong K. Choi, Alyssa Bosold, Katrin Kirchhoff	Journal of Biomedical Informatics	University of Washington	Estados Unidos	2018	Si
95	Science direct	Future of evidence ecosystem series: 2. current opportunities and need for better tools and methods	To become user driven and more useful for decision-making, the current evidence synthesis ecosystem requires significant changes (Paper 1. Future of evidence ecosystem series). Reviewers have access to new sources of data (clinical trial registries, protocols, and clinical study reports from regulatory agencies or pharmaceutical companies) for more information on randomized control trials. With all these newly available data, the management of multiple and scattered trial reports is even more challenging. New types of data are also becoming available: individual patient data and routinely collected data. With the increasing number of diverse sources to be searched and the amount of data to be extracted, the process needs to be rethought. New approaches and tools, such as automation technologies and crowdsourcing, should help accelerate the process. The implementation of these new approaches and methods requires a substantial rethinking and redesign of the current evidence synthesis ecosystem. The concept of a “living” evidence synthesis enterprise, with living systematic review and living network meta-analysis, has recently emerged. Such an evidence synthesis ecosystem implies conceptualizing evidence synthesis as a continuous process built around a clinical question of interest and no longer as a small team independently answering a specific clinical question at a single point in time.	PerrineCréquit, IsabelleBoutron, JoergMeerpohl, Hywel C.Williams, JonathanCraig, PhilippeRavaud	Journal of Clinical Epidemiology	Paris Descartes University, University of Nottingham, Flinders University, Columbia University, University of Freiburg	Francia, Alemania, Reino Unido, Australia, Estados Unidos	2020	Si
96	Science direct	Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals	While systematic review (SR) methods are gaining traction as a method for providing a reliable summary of existing evidence for health risks posed by exposure to chemical substances, it is becoming clear that their value is restricted to a specific range of risk management scenarios - in particular, those which can be addressed	Taylor A.M.Wolffe, PaulWhaley,	Environment International	Lancaster University,	Estados Unidos, Reino Unido	2019	Si

		policy and risk management	with tightly focused questions and can accommodate the time and resource requirements of a systematic evidence synthesis.	CrispinHalsall, Andrew A.Rooney, Vickie R.Walker					
97	Science direct	Automatic extraction of quantitative data from ClinicalTrials.gov to conduct meta-analyses	Systematic reviews and meta-analyses are labor-intensive and time-consuming. Automated extraction of quantitative data from primary studies can accelerate this process. ClinicalTrials.gov, launched in 2000, is the world's largest trial repository of results data from clinical trials; it has been used as a source instead of journal articles. We have developed a Web application called EXACT (EXtracting Accurate efficacy and safety information from ClinicalTrials.gov) that allows users without advanced programming skills to automatically extract data from ClinicalTrials.gov in analysis-ready format. We have also used the automatically extracted data to examine the reproducibility of meta-analyses in three published systematic reviews.	RicheekPradhan, David C.Hoaglin, MatthewCornell, WeisongLiu, VictoriaWang, HongYu	Journal of Clinical Epidemiology	University of Massachusetts	Estados Unidos	2019	Si
98	Science direct	Future of Evidence Ecosystem Series: Evidence synthesis 2.0: when systematic, scoping, rapid, living, and overviews of reviews come together		Elie A.Akl, Neal R.Haddaway, GabrielRada, TamaraLotfi	Journal of Clinical Epidemiology	American University of Beirut, McMaster University, University of Johannesburg, Pontificia Universidad Católica de Chile	Libano, Canada, Suecia, Sudafrica, Chile	2020	Si
99	Science direct	Three-dimensional acquisition technologies for facial soft tissues – Applications and prospects in orthognathic surgery	The management of patients with dento-maxillofacial deformities is based on assessments of the dental occlusion – facial skeleton – soft tissues triad. As societal demands and surgical practices have evolved, facial soft tissues have moved to the forefront of considerations in orthognathic surgery. Techniques are therefore required to analyze facial soft tissues objectively and reproducibly, for diagnosis, preoperative planning, and follow-up. Several technologies are currently capable of providing three-dimensional (3D) models of the face, either by 3D reconstruction of traditional computed tomography or cone beam computed tomography data, or directly by stereophotogrammetry, laser scanning or structured light scanning. Multimodal image registration techniques allow bone base, dental occlusion and facial soft tissue information to be combined in a 3D virtual patient. Three-dimensional cephalometric analysis of the facial skeleton and skin is now perfectly integrated in virtual planning and is gradually gaining in automation and accuracy. Photorealistic 3D simulations allow optimal soft tissue planning and facilitate physician-patient communication. Finally, these facial modeling techniques facilitate post-operative studies of soft tissues, which generally involve comparisons of volumetric data. There are many research avenues to pursue and technical improvements are to be	S.Rasteau, N.Sigaux, A.Louvrier, P.Bouletreau	Journal of Stomatology, Oral and Maxillofacial Surgery	université Claude-Bernard Lyon,	Francia	2020	Si

			expected, particularly through the development of big data and artificial intelligence approaches.						
100	Science direct	Critical Analysis of Artificial Intelligence in Endodontics: A Scoping Review	Artificial intelligence (AI) comprises computational models that mimic the human brain to perform various diagnostic tasks in clinical practice. The aim of this scoping review was to systematically analyze the AI algorithms and models used in endodontics and identify the source quality and type of evidence.	FahadUmer, SaqibHabib	Journal of Endodontics	FahadUmer, SaqibHabib	Pakistan	2022	Si
101	Science direct	Advanced vehicle technologies and road safety: A scoping review of the evidence	The proliferation of Advanced Vehicle Technologies (AVTs) has generated both excitement and concern among researchers, policymakers, and the general public. An increasing number of driver assistance systems are already available in today's automobiles; many of which are expected to become standard. Therefore, synthesizing the available evidence specific to the safety of AVTs is critical. The goal of this scoping review was to summarize this evidence with a focus on AVTs that require some driver oversight (i.e., Levels 0–3 as per the Society of Automotive Engineers (SAE) levels of automation taxonomy). A scoping review of research literature on AVTs was conducted for studies up to March 2018. Inclusion criteria consisted of: any study with empirical data of AVTs that included male and female drivers aged 16 years and older, healthy people (i.e., without impairments), passenger vehicles, driving simulators and/or large databases with road safety information that could be analyzed for the purpose of examining AVTs (SAE Levels 0–3), as well as measures of driving outcomes. A total of 324 peer-reviewed studies from 25 countries met the inclusion criteria for this review with over half published in the last 5 years. Data was extracted and summarized according to the following categories: measures used to evaluate the effect of AVTs on road safety (objective) and driver perceptions of the technology (subjective), testing environment, and study populations (i.e., driver age). The most commonly reported objective measures were longitudinal control (50 %), reaction time (40 %), and lateral position (23 %). The most common subjective measures were perceptions of trust (27 %), workload (20 %), and satisfaction (17 %). While most studies investigated singular AVTs (237 of 324 studies), the number of studies after 2013 that examined 2 or more AVTs concurrently increased. Studies involved drivers from different age groups (51 %) and were conducted in driving simulators (70 %). Overall, the evidence is generally in favour of AVTs having a positive effect on	Andrea D.Furlan, TaraKajaks, MargaretTiong, MartinLavallière, Jennifer L.Campos, JessicaBabineau, Shabnam Haghzare, TraceyMa, BrendaVrkljan	Accident Analysis & Prevention	University of Toronto, University of New South Wales	Canada, Australia	2020	Si

			driving safety, although the nature and design of studies varied widely. Our examination of this evidence highlights the opportunities as well as the challenges involved with investigating AVTs. Ensuring such technologies are congruent with the needs of drivers, particularly younger and older driver age groups, who are known to have a higher crash risk, is critical. With automotive manufacturers keen to adopt the latest AVTs, this scoping review highlights how testing of this technology has been undertaken, with a focus on how new research can be conducted to improve road safety now and in the future.						
102	Science direct	Maintenance transformation through Industry 4.0 technologies: A systematic literature review	Industry 4.0 is revolutionizing manufacturing, increasing flexibility, mass customization, quality and productivity. In today's competitive manufacturing scenario, maintenance is one of the most critical issues and companies are approaching its digital transformation from technological and management perspectives. This article carries out a systematic literature review aimed to investigate how maintenance tasks and maintenance management strategies are changing in Industry 4.0 context, analyzing the state-of-the-art of Industry 4.0 technologies currently employed in maintenance and the resulting potential innovations in maintenance policies and manufacturing management. In addition, the most relevant trends in current maintenance policies have been investigated, such as “remote maintenance” and the attractive possibility of a “self-maintenance”. Also, the importance of human factor has been considered. The results are summarized in a comprehensive database, to provide, through concepts and empirical evidence present in literature, examples and strategies for the implementation of maintenance in Industry 4.0.	LucaSilvestri, AntonioForcina, VitoIntrona, AnnalisaSantolamazza, VittorioCesarotti	Computers in Industry	Niccolò Cusano University, University of Naples “Parthenope”, “Tor Vergata” University of Rome	Italia	2020	Si
103	Science direct	Computational methods and rural cultural & natural heritage: A review	Cultural and Natural Heritage (CNH) are both irreplaceable sources of life and inspiration, according to the UNESCO definition. Rural areas represent outstanding examples of cultural, either tangible or intangible, and natural heritage. While rural areas are facing a socio-economic and demographic crisis all over the world, CNH need not only to be safeguarded, but also promoted as a driver for competitiveness, growth and sustainable and inclusive development. This paper goes deeper into the study of computational methods (CMs) applied to modelling CNH in rural areas by looking at how computational methods can support CNH promotion and valorisation to transform rural areas into laboratories for the demonstration of sustainable development through improving the unique potential of	Francisco Barrientos, JohnMartin, ClaudiaDeLuca, SimonaTondelli, JaimeGómez-García-Bermejo, Eduardo	Journal of Cultural Heritage	University of Plymouth, University of Bologna, University of Valladolid	España, Italia, Reino Unido	2021	Si

			their heritage. To this end, different computational methods have been studied and classified according to their scope and application area parameters, showing some correlation among the said parameters and the class of computational method. Apart from how CMs have been applied, wehether it is possible to scale up these CMs elsewhere has also been considered.	ZalamaCasanova					
104	Science direct	The four dimensions of contestable AI diagnostics - A patient-centric approach to explainable AI	The problem of the explainability of AI decision-making has attracted considerable attention in recent years. In considering AI diagnostics we suggest that explainability should be explicated as 'effective contestability'. Taking a patient-centric approach we argue that patients should be able to contest the diagnoses of AI diagnostic systems, and that effective contestation of patient-relevant aspect of AI diagnoses requires the availability of different types of information about 1) the AI system's use of data, 2) the system's potential biases, 3) the system performance, and 4) the division of labour between the system and health care professionals. We justify and define thirteen specific informational requirements that follows from 'contestability'. We further show not only that contestability is a weaker requirement than some of the proposed criteria of explainability, but also that it does not introduce poorly grounded double standards for AI and health care professionals' diagnostics, and does not come at the cost of AI system performance. Finally, we briefly discuss whether the contestability requirements introduced here are domain-specific.	ThomasPloug, SørenHolm	Artificial Intelligence in Medicine	Aalborg University, University of Manchester, University of Oslo	Dinamarca, Noruega, Reino Unido	2020	Si
105	Science direct	Constructing public health evidence knowledge graph for decision-making support from COVID-19 literature of modelling study	The needs of mitigating COVID-19 epidemic prompt policymakers to make public health-related decision under the guidelines of science. Tremendous unstructured COVID-19 publications make it challenging for policymakers to obtain relevant evidence. Knowledge graphs (KGs) can formalize unstructured knowledge into structured form and have been used in supporting decision-making recently. Here, we introduce a novel framework that can extract the COVID-19 public health evidence knowledge graph (CPHE-KG) from papers relating to a modelling study. We screen out a corpus of 3096 COVID-19 modelling study papers by performing a literature assessment process. We define a novel annotation schema to construct the COVID-19 modelling study-related IE dataset (CPHIE). We also propose a novel multi-tasks document-level information extraction model SS-DYGIE++ based on the dataset. Leveraging the model on the new corpus, we construct CPHE-KG containing 60,967 entities and 51,140 relations. Finally, we seek to apply our KG to support evidence querying and evidence mapping visualization. Our SS-DYGIE++(SpanBERT) model has achieved a F1 score of 0.77 and 0.55 respectively in document-level entity recognition and coreference resolution tasks. It has also shown high performance in the relation identification task. With evidence querying, our KG can present the dynamic transmissions of COVID-19 pandemic in different countries and regions. The evidence mapping of our KG	YunrongYang, ZhidongCao, PengfeiZhao, Dajun DanielZeng, Qingpeng Zhang, YinLuo	Journal of Safety Science and Resilience	University of Chinese Academy of Sciences, City University of Hong Kong - Hong Kong SAR	China	2021	Si

			can show the impacts of variable non-pharmacological interventions to COVID-19 pandemic. Analysis demonstrates the quality of our KG and shows that it has the potential to support COVID-19 policy making in public health.						
106	Science direct	The Role of Risk Prediction Models in Prevention and Management of AKI	Acute kidney injury is a major health care problem. Improving recognition of those at risk and highlighting those who have developed AKI at an earlier stage remains a priority for research and clinical practice. Prediction models to risk-stratify patients and electronic alerts for AKI are two approaches that could address previously highlighted shortcomings in management and facilitate timely intervention. We describe and critique available prediction models and the effects of the use of AKI alerts on patient outcomes are reviewed. Finally, the potential for prediction models to enrich population subsets for other diagnostic approaches and potential research, including biomarkers of AKI, are discussed.	Luke E.Hodgson, NicholasSelby, Tao-MinHuang, Lui G.Forni	Seminars in Nephrology			2019	Disponible en clinical Key
107	Science direct	Buhos: A web-based systematic literature review management software	Software can significantly facilitate the management of the complete systematic literature review process (SLR). However, most specialized software for use in SLR processes is designed to meet the requirements of the health and medical sciences and software engineering, and there is a need for dedicated software for the specific research requirements of the social sciences. Furthermore, most of the software currently used is closed and the open source code alternatives require personnel with expertise in configuration and setup. We present Buhos, an application for managing the complete process of systematic literature reviews that is web-based and developed in Ruby. It offers functionalities for supporting the process of searching, screening, data extraction and reporting. Buhos can be used locally through an in-house web server, as well as in a distributed manner, integrated with other online services.	ClaudioBustos Navarrete, María GabrielaMorales Malverde, PedroSalcedo Lagos, Alejandro Díaz Mujica	SoftwareX	Universidad de Concepción,	Chile	2018	Si
108	Science direct	The evolution of robotic surgery: surgical and anaesthetic aspects	Robotic surgery pushes the frontiers of innovation in healthcare technology towards improved clinical outcomes. We discuss the evolution to five generations of robotic surgical platforms including stereotactic, endoscopic, bioinspired, microbots on the millimetre scale, and the future development of autonomous systems. We examine the challenges, obstacles and limitations of robotic surgery and its future potential including integrated real-time anatomical and immune-histological imaging and data assimilation with improved visualisation, haptic feedback and robot-surgeon interactivity. We consider current evidence, cost-effectiveness and the learning curve in relation to the surgical and anaesthetic journey, and what is required to continue to realise improvements in surgical operative	H.Ashrafian, O.Clancy, V.Grover, A.Darzi	British Journal of Anaesthesia	Imperial College London	Reino Unido	2017	Si

			care. The innovative impact of this technology holds the potential to achieve transformative clinical improvements. However, despite over 30 yr of incremental advances it remains formative in its innovative disruption.						
109	Science direct	Risk management in sustainable smart cities governance: A TOE framework	Sustainable smart cities are confronted by technological, organisational and external risks, making their governance difficult and susceptible to manipulation. Based on a comprehensive literature review of 796 systematically retrieved articles, the current study proposes a multilayered technology-organisation-environment (TOE-based) risk management framework for sustainable smart city governance. A total of 56 risks are identified and grouped into TOE categories. There are 17 technological risks, including IoT networks, public internet management and user safety concerns, with a 38.7% contribution to smart city governance risks. With a 15.6% share, there are 11 organisational risks, including user data security and cloud management. There are 28 external risks with a contribution of 46.7% to the smart city governance and consist of smart city's environment, governance, integration and security risks. A multilayered TOE-based risk management framework is proposed to identify and manage the risks associated with smart city governance in the current study. The framework links smart citizens to each other through the smart city governance team and the integrated TOE layers. The iterative risk management process of identification, analysis, evaluation, monitoring and response planning is carried out in the TOE layers, both at the external layer levels and internal management levels. The proposed framework operationalises the risk management process for smart city governance by presenting the collection of pertinent risks and their thematic TOE categorisation. The criticality of the identified risks in line with the study's rankings can help researchers and practitioners understand the top risks of smart city governance. These risks present investment opportunities for city governance bodies to develop critical and effective responses as well as provide safety, security and enhanced privacy for citizens.	FahimUllah, SiddraQayyum, Muhammad JamaluddinThaheem	Technological Forecasting and Social Change	University of New South Wales (UNSW) Sydney, Deakin University, Near East University	Australia, Turkia	2021	Si
110	Science direct	AI-based monitoring of retinal fluid in disease activity and under therapy	Retinal fluid as the major biomarker in exudative macular disease is accurately visualized by high-resolution three-dimensional optical coherence tomography (OCT), which is used world-wide as a diagnostic gold standard largely replacing clinical examination. Artificial intelligence (AI) with its capability to objectively identify, localize and quantify fluid introduces fully automated tools into OCT imaging for personalized disease management. Deep learning performance has already proven superior to human experts, including physicians and certified readers, in terms of accuracy and speed. Reproducible measurement of retinal fluid relies on precise AI-based segmentation methods that assign a label to each OCT voxel denoting its fluid type such as intraretinal fluid (IRF) and	UrsulaSchmidt-Erfurth, Gregor S.Reiter, SophieRiedl	Progress in Retinal and Eye Research	University of Vienna, University of Wisconsin, Northwestern University, Oregon Health & Science University, University of California Los Angeles	Estados Unidos, Austria	2022	Si

			subretinal fluid (SRF) or pigment epithelial detachment (PED) and its location within the central 1-, 3- and 6-mm macular area. Such reliable analysis is most relevant to reflect differences in pathophysiological mechanisms and impacts on retinal function, and the dynamics of fluid resolution during therapy with different regimens and substances. Yet, an in-depth understanding of the mode of action of supervised and unsupervised learning, the functionality of a convolutional neural net (CNN) and various network architectures is needed. Greater insight regarding adequate methods for performance, validation assessment, and device- and scanning-pattern-dependent variations is necessary to empower ophthalmologists to become qualified AI users. Fluid/function correlation can lead to a better definition of valid fluid variables relevant for optimal outcomes on an individual and a population level. AI-based fluid analysis opens the way for precision medicine in real-world practice of the leading retinal diseases of modern times.						
111	Science direct	Abstracts from The International Symposium on Endovascular Therapy (ISET) 2019			Journal of Vascular and Interventional Radiology			2019	Disponible en clinical Key
112	Springer Link	Machine Learning in Dentistry		Ching-Chang KoDinggang ShenLi Wang				2021	Si
113	Springer Link	Principles and Practice of Systematic Reviews and Meta-Analysis		Sanjay Patole				2021	Si
114	Springer Link	Experimental Design and Reproducibility in Preclinical Animal Studies		José M. Sánchez MorgadoAurora Brønstad				2021	Si
115	Springer Link	Access to Non-Summary Clinical Trial Data for Research Purposes Under EU Law		Daria Kim				2021	No
116	Springer Link	Natural Products and Human Diseases		Amirhossein Sahebkar Thozhukat Sathyapalan				2021	No
117	Springer Link	Machine learning to help researchers	Assessing risks of bias in randomized controlled trials (RCTs) is an important but laborious task when conducting systematic reviews.	Frank Soboczen	BMC Medical Informatics			2019	Si

		evaluate biases in clinical trials: a prospective, randomized user study	RobotReviewer (RR), an open-source machine learning (ML) system, semi-automates bias assessments. We conducted a user study of RobotReviewer, evaluating time saved and usability of the tool.	ski, Thomas A. Trikalinos, Joël Kuiper	and Decision Making				
118		A systematic review of the applications of Expert Systems (ES) and machine learning (ML) in clinical urology	Testing a hypothesis for 'factors-outcome effect' is a common quest, but standard statistical regression analysis tools are rendered ineffective by data contaminated with too many noisy variables. Expert Systems (ES) can provide an alternative methodology in analysing data to identify variables with the highest correlation to the outcome. By applying their effective machine learning (ML) abilities, significant research time and costs can be saved. The study aims to systematically review the applications of ES in urological research and their methodological models for effective multi-variate analysis. Their domains, development and validity will be identified.	Hesham Salem, Daniele Soria, Jonathan N. Lund & Amir Awwad	BMC Medical Informatics and Decision Making			2021	Si
119	Springer Link	Thoracic imaging radiomics for staging lung cancer: a systematic review and radiomic quality assessment	Radiomics, a method used to extract large amount of data, may be useful for discriminating malignant characteristics for lung cancer staging. We aimed to critically appraise the current use of radiomics in medical imaging for determining nodal involvement in lung cancer.	Isabella F. Churchill, Kerrie A. Sullivan, Alexander C. Simone, Yogita S. Patel, Grigorios I. Leontiadis,				2021	Si
120	Springer Link	Can minimal clinically important differences in patient reported outcome measures be predicted by machine learning in patients with total knee or hip arthroplasty? A systematic review	To systematically review studies using machine learning (ML) algorithms to predict whether patients undergoing total knee or total hip arthroplasty achieve an improvement as high or higher than the minimal clinically important differences (MCID) in patient reported outcome measures (PROMs) (classification problem).	Benedikt Langenberger, Andreas Thoma & Verena Vogt	BMC Medical Informatics and Decision Making			2022	Si
121	Springer Link	Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy	Early clinical recognition of sepsis can be challenging. With the advancement of machine learning, promising real-time models to predict sepsis have emerged. We assessed their performance by carrying out a systematic review and meta-analysis.	Lucas M. Fleuren, Thomas L. T. Klausch, Charlotte L. Zwager	Intensive Care Medicine			2020	Si
122	Springer Link	Development, validation and effectiveness of	Tools based on diagnostic prediction models are available to help general practitioners (GP) diagnose colorectal cancer. It is unclear how well they perform and whether they lead to increased or quicker	Bogdan Grigore, Ruth	BMC Cancer			2020	Si

		diagnostic prediction tools for colorectal cancer in primary care: a systematic review|	diagnoses and ultimately impact on patient quality of life and/or survival. The aim of this systematic review is to evaluate the development, validation, effectiveness, and cost-effectiveness, of cancer diagnostic tools for colorectal cancer in primary care.	Lewis, Jaime Peters, Sophie Robinson & Christopher J. Hyde					
123	Springer Link	Prognostic scores for survival as decisional support for surgery in spinal metastases: a performance assessment systematic review	To review the evidence on the relative prognostic performance of the available prognostic scores for survival in spinal metastatic surgery in order to provide a recommendation for use in clinical practice.	S. Smeijers & B. Depreitere	European Spine			2021	Si
124	Springer Link	Prognostic models for amyotrophic lateral sclerosis: a systematic review	Increasing prognostic models for amyotrophic lateral sclerosis (ALS) have been developed. However, no comprehensive evaluation of these models has been done. The purpose of this study was to map the prognostic models for ALS to assess their potential contribution and suggest future improvements on modeling strategy.	Lu Xu, Bingjie He, Yunjing Zhang, Lu Chen, Dongsheng Fan, Siyan Zhan, & Shengfeng Wang	Journal of Neurology			2021	Si
125	Springer Link	Methodological quality of machine learning-based quantitative imaging analysis studies in esophageal cancer: a systematic review of clinical outcome prediction after concurrent chemoradiotherapy	Studies based on machine learning-based quantitative imaging techniques have gained much interest in cancer research. The aim of this review is to critically appraise the existing machine learning-based quantitative imaging analysis studies predicting outcomes of esophageal cancer after concurrent chemoradiotherapy in accordance with PRISMA guidelines.	Zhenwei Shi, Zhen Zhang, Zaiyi Liu, Lujun Zhao, Zhaoxiang Ye, Andre Dekker & Leonard Wee	European Journal of Nuclear Medicine and Molecular Imaging			2021	Si
126	Springer Link	Toward systematic review automation: a practical guide to using machine learning tools in research synthesis	Technologies and methods to speed up the production of systematic reviews by reducing the manual labour involved have recently emerged. Automation has been proposed or used to expedite most steps of the systematic review process, including search, screening, and data extraction. However, how these technologies work in practice and when (and when not) to use them is often not clear to practitioners. In this practical guide, we provide an overview of current machine learning methods that have been proposed to expedite evidence synthesis. We also offer guidance on which of these are ready for use, their strengths and weaknesses, and how a systematic review team might go about using them in practice.	Iain J. Marshall & Byron C. Wallace	Systematic Reviews			2019	Si

127	Springer Link	Machine Learning in Evidence Synthesis Research	In this chapter, we will explore how Systemic Reviews (SR) are traditionally conducted and how the process of arriving at a valuable SR can be made more efficient and less error prone using Machine Learning (ML) techniques. As the integration of ML at the screening stage of SRs has reached the highest level of maturity, we will explain the techniques utilized. We will further describe the extraction process from primary studies supported by ML techniques. The discussion of pitfalls when conducting SRs concludes the chapter, specifically how ML can address bias. Lastly, we address the inherent limitations of artificial intelligence in healthcare with a special emphasis on ML for the use in SRs.	Alonso Carrasco-Labra				2021	Si
128	Springer Link	Survival prediction of glioblastoma patients—are we there yet? A systematic review of prognostic modeling for glioblastoma and its clinical potential	Glioblastoma is associated with a poor prognosis. Even though survival statistics are well-described at the population level, it remains challenging to predict the prognosis of an individual patient despite the increasing number of prognostic models. The aim of this study is to systematically review the literature on prognostic modeling in glioblastoma patients. A systematic literature search was performed to identify all relevant studies that developed a prognostic model for predicting overall survival in glioblastoma patients following the PRISMA guidelines. Participants, type of input, algorithm type, validation, and testing procedures were reviewed per prognostic model. Among 595 citations, 27 studies were included for qualitative review. The included studies developed and evaluated a total of 59 models, of which only seven were externally validated in a different patient cohort. The predictive performance among these studies varied widely according to the AUC (0.58–0.98), accuracy (0.69–0.98), and C-index (0.66–0.70). Three studies deployed their model as an online prediction tool, all of which were based on a statistical algorithm. The increasing performance of survival prediction models will aid personalized clinical decision-making in glioblastoma patients. The scientific realm is gravitating towards the use of machine learning models developed on high-dimensional data, often with promising results. However, none of these models has been implemented into clinical care. To facilitate the clinical implementation of high-performing survival prediction models, future efforts should focus on harmonizing data acquisition methods, improving model interpretability, and externally validating these models in multicentered, prospective fashion.	Ishaan Ashwini Tewarie, Joeky T. Senders, Stijn Kremer, Sharmila Devi, William B. Gormley, Omar Arnaout, Timothy R. Smith & Marike L. D. Broekman	Neurosurgical Review			2021	Si
129	Springer Link	Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit	Due to the increasing demand for intensive care unit (ICU) treatment, and to improve quality and efficiency of care, there is a need for adequate and efficient clinical decision-making. The advancement of artificial intelligence (AI) technologies has resulted in the development of prediction models, which might aid clinical decision-making. This systematic review seeks to give a contemporary overview of the current maturity of AI in the ICU, the research methods behind these studies, and the risk of bias in these studies.	Davy van de Sande, Michel E. van Genderen, Joost Huiskens, Diederik Gommers & Jasper	Intensive Care Medicine			2021	Si

				van Bommel					
130	Springer Link	The applications of machine learning in plastic and reconstructive surgery: protocol of a systematic review	Machine learning, a subset of artificial intelligence, is a set of models and methods that can automatically detect patterns in vast amounts of data, extract information and use it to perform various kinds of decision-making under uncertain conditions. This can assist surgeons in clinical decision-making by identifying patient cohorts that will benefit from surgery prior to treatment. The aim of this review is to evaluate the applications of machine learning in plastic and reconstructive surgery.	Angelos Mantelakis & Ankur Khajuria	Systematic Reviews			2020	Si
131	Springer Link	Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature	Clinical prediction rules (CPRs) that predict the absolute risk of a clinical condition or future outcome for individual patients are abundant in the medical literature; however, systematic reviews have demonstrated shortcomings in the methodological quality and reporting of prediction studies. To maximise the potential and clinical usefulness of CPRs, they must be rigorously developed and validated, and their impact on clinical practice and patient outcomes must be evaluated. This review aims to present a comprehensive overview of the stages involved in the development, validation and evaluation of CPRs, and to describe in detail the methodological standards required at each stage, illustrated with examples where appropriate. Important features of the study design, statistical analysis, modelling strategy, data collection, performance assessment, CPR presentation and reporting are discussed, in addition to other, often overlooked aspects such as the acceptability, cost-effectiveness and longer-term implementation of CPRs, and their comparison with clinical judgement. Although the development and evaluation of a robust, clinically useful CPR is anything but straightforward, adherence to the plethora of methodological standards, recommendations and frameworks at each stage will assist in the development of a rigorous CPR that has the potential to contribute usefully to clinical practice and decision-making and have a positive impact on patient care.	Laura E. Cowley, Daniel M. Farewell, Sabine Maguire & Alison M. Kemp	Diagnostic and Prognostic Research			2019	Si
132	Springer Link	Machine Learning and Glioblastoma: Treatment Response Monitoring Biomarkers in 2021	The aim of the systematic review was to assess recently published studies on diagnostic test accuracy of glioblastoma treatment response monitoring biomarkers in adults, developed through machine learning (ML). Articles published 09/2018–09/2020 were searched for using MEDLINE, EMBASE, and the Cochrane Register. Included study participants were adult patients with high grade glioma who had undergone standard treatment (maximal resection, radiotherapy with concomitant and adjuvant temozolomide) and subsequently underwent follow-up imaging to determine treatment response status (specifically, distinguishing progression/recurrence from progression/recurrence mimics - the target condition). Risk of bias and applicability was assessed with QUADAS 2 methodology. Contingency tables were created for hold-out test sets and recall,	Thomas C. BoothEmail authorBernice AkpinarAndrei RomanHaris ShuaibAysha LuisAlysha				2020	Si

			specificity, precision, F1-score, balanced accuracy calculated. Fifteen studies were included with 1038 patients in training sets and 233 in test sets. To determine whether there was progression or a mimic, the reference standard combination of follow-up imaging and histopathology at re-operation was applied in 67% (10/15) of studies. External hold-out test sets were used in 27% (4/15) to give ranges of diagnostic accuracy measures: recall = 0.70–1.00; specificity = 0.67–0.90; precision = 0.78–0.88; F1 score = 0.74–0.94; balanced accuracy = 0.74–0.83; AUC = 0.80–0.85. The small numbers of patient included in studies, the high risk of bias and concerns of applicability in the study designs (particularly in relation to the reference standard and patient selection due to confounding), and the low level of evidence, suggest that limited conclusions can be drawn from the data. There is likely good diagnostic performance of machine learning models that use MRI features to distinguish between progression and mimics. The diagnostic performance of ML using implicit features did not appear to be superior to ML using explicit features. There are a range of ML-based solutions poised to become treatment response monitoring biomarkers for glioblastoma. To achieve this, the development and validation of ML models require large, well-annotated datasets where the potential for confounding in the study design has been carefully considered. Therefore, multidisciplinary efforts and multicentre collaborations are necessary.	ChelliahAyisha Al BusaidiAyesha MirchandaniBurcu AlparslanNina MansoorKeyoumarsAshkanSebastien OurselinMarc Modat					
133	Springer Link	Experimental Design and Reproducibility in Preclinical Animal Studies	Systematic reviews are a firmly established method of ensuring that proposed research is based upon the best available scientific evidence. In this chapter, we provide a brief history of systematic reviews and discuss their adaptation to preclinical studies. The steps in conducting a systematic review are explained, with examples of best practice. Readers will learn how to critically evaluate the quality of systematic reviews in their own fields. Basic guidance on the parts of a systematic review and meta-analysis are explained. Critically appraised topics (or knowledge summaries) are also described, and their relevance for preclinical research is explained, including a worked example.	Janet Becker RodgersEmail authorMerel Ritskes-Hoitinga				2021	Si
134	Springer Link	Automated COVID-19 diagnosis and prognosis with medical imaging and who is publishing: a systematic review	Objectives: To conduct a systematic survey of published techniques for automated diagnosis and prognosis of COVID-19 diseases using medical imaging, assessing the validity of reported performance and investigating the proposed clinical use-case. To conduct a scoping review into the authors publishing such work. Methods: The Scopus database was queried and studies were screened for article type, and minimum source normalized impact per paper and citations, before manual relevance assessment and a bias assessment derived from a subset of the Checklist for Artificial Intelligence in Medical Imaging (CLAIM). The number of failures of the full CLAIM was adopted as a surrogate for risk-of-bias. Methodological and performance measurements were collected from each technique. Each study was assessed by one author. Comparisons were	Ashley G. Gillman, Febrio Lunardo, Joseph Prinable, Gregg Belous, Aaron Nicolson, Hang Min, Andrew Terhorst &	Physical and Engineering Sciences in Medicine			2021	Si

			evaluated for significance with a two-sided independent t-test. Findings: Of 1002 studies identified, 390 remained after screening and 81 after relevance and bias exclusion. The ratio of exclusion for bias was 71%, indicative of a high level of bias in the field. The mean number of CLAIM failures per study was 8.3 ± 3.9 [1,17] (mean ± standard deviation [min,max]). 58% of methods performed diagnosis versus 31% prognosis. Of the diagnostic methods, 38% differentiated COVID-19 from healthy controls. For diagnostic techniques, area under the receiver operating curve (AUC) = 0.924 ± 0.074 [0.810,0.991] and accuracy = 91.7% ± 6.4 [79.0,99.0]. For prognostic techniques, AUC = 0.836 ± 0.126 [0.605,0.980] and accuracy = 78.4% ± 9.4 [62.5,98.0]. CLAIM failures did not correlate with performance, providing confidence that the highest results were not driven by biased papers. Deep learning techniques reported higher AUC ($p < 0.05$) and accuracy ($p < 0.05$), but no difference in CLAIM failures was identified. Interpretation: A majority of papers focus on the less clinically impactful diagnosis task, contrasted with prognosis, with a significant portion performing a clinically unnecessary task of differentiating COVID-19 from healthy. Authors should consider the clinical scenario in which their work would be deployed when developing techniques. Nevertheless, studies report superb performance in a potentially impactful application. Future work is warranted in translating techniques into clinical tools.	Jason A. Dowling					
135	Springer Link	Automating data extraction in systematic reviews: a systematic review	Automation of the parts of systematic review process, specifically the data extraction step, may be an important strategy to reduce the time necessary to complete a systematic review. However, the state of the science of automatically extracting data elements from full texts has not been well described. This paper performs a systematic review of published and unpublished methods to automate data extraction for systematic reviews.	Siddhartha R. Jonnalagadda, Pawan Goyal & Mark D. Huffman	Systematic Reviews			2015	Si
136	Springer Link	Lung ultrasound training: a systematic review of published literature in clinical lung ultrasound training	Clinical lung ultrasound examinations are widely used in the primary assessment or monitoring of patients with dyspnoea or respiratory failure. Despite being increasingly implemented, there is no international consensus on education, assessment of competencies, and certification. Today, training is usually based on the concept of mastery learning, but is often unstructured and limited by bustle in a clinical daily life. The aim of the systematic review is to provide an overview of published learning studies in clinical lung ultrasound, and to collect evidence for future recommendations in lung ultrasound education and certification.	Pia Iben Pietersen, Kristian Rørbæk Madsen, Ole Graumann, Lars Konge, Bjørn Ulrik Nielsen & Christian Borbjerg Laursen	Critical Ultrasound			2018	Si

137	Springer Link	A systematic review of methodology used in the development of prediction models for future asthma exacerbation	Clinical prediction models are widely used to guide medical advice and therapeutic interventions. Asthma is one of the most common chronic diseases globally and is characterised by acute deteriorations. These exacerbations are largely preventable, so there is interest in using clinical prediction models in this area. The objective of this review was to identify studies which have developed such models, determine whether consistent and appropriate methodology was used and whether statistically reliable prognostic models exist.	Joshua Bridge, John D. Blakey & Laura J. Bonnett	BMC Medical Research Methodology			2020	Si
138	Springer Link	Effectiveness of technology-enhanced teaching and assessment methods of undergraduate preclinical dental skills: a systematic review of randomized controlled clinical trials	To investigate the effectiveness of technology-enhanced teaching and assessment methods of undergraduate preclinical skills in comparison to conventional methods.	Khaled Khalaf, Mohamed El-Kishawi, Shahd Mustafa & Sausan Al Kawas	BMC Medical Education			2020	Si
139	Springer Link	Application of artificial intelligence and radiomics in pituitary neuroendocrine and sellar tumors: a quantitative and qualitative synthesis	To systematically review the literature regarding the application of machine learning (ML) of magnetic resonance imaging (MRI) radiomics in common sellar tumors. To identify future directions for application of ML in sellar tumor MRI.	Kelvin Koong, Veronica Preda, Anne Jian, Benoit Liquet-Weiland & Antonio Di Ieva	Neuroradiology			2021	Si
140	Springer Link	A systematic review and meta-analysis of online versus alternative methods for training licensed health care professionals to deliver clinical interventions	Online training is growing in popularity and yet its effectiveness for training licensed health professionals (HCPs) in clinical interventions is not clear. We aimed to systematically review the literature on the effectiveness of online versus alternative training methods in clinical interventions for licensed Health Care Professionals (HCPs) on outcomes of knowledge acquisition, practical skills, clinical behaviour, self-efficacy and satisfaction.	Helen Richmond, Bethan Copsey, Amanda M. Hall, David Davies & Sarah E. Lamb	BMC Medical Education			2017	Si
141	Springer Link	Machine Learning in Medical Emergencies: a Systematic Review and Analysis	Despite the increasing demand for artificial intelligence research in medicine, the functionalities of his methods in health emergency remain unclear. Therefore, the authors have conducted this systematic review and a global overview study which aims to identify, analyse, and evaluate the research available on different platforms, and its implementations in healthcare emergencies. The methodology applied for the identification and selection of the scientific studies and the different applications consist of two methods. On the one hand, the PRISMA methodology was carried	Inés Robles Mendo, Gonçalo Marques, Isabel de la Torre Díez, Miguel	Journal of Medical Systems			2021	Si

			out in Google Scholar, IEEE Xplore, PubMed ScienceDirect, and Scopus. On the other hand, a review of commercial applications found in the best-known commercial platforms (Android and iOS). A total of 20 studies were included in this review. Most of the included studies were of clinical decisions (n = 4, 20%) or medical services or emergency services (n = 4, 20%). Only 2 were focused on m-health (n = 2, 10%). On the other hand, 12 apps were chosen for full testing on different devices. These apps dealt with pre-hospital medical care (n = 3, 25%) or clinical decision support (n = 3, 25%). In total, half of these apps are based on machine learning based on natural language processing. Machine learning is increasingly applicable to healthcare and offers solutions to improve the efficiency and quality of healthcare. With the emergence of mobile health devices and applications that can use data and assess a patient's real-time health, machine learning is a growing trend in the healthcare industry.	López-Coronado & Francisco Martín-Rodríguez					
142	Springer Link	The effects of error-augmentation versus error-reduction paradigms in robotic therapy to enhance upper extremity performance and recovery post-stroke: a systematic review	Despite upper extremity function playing a crucial role in maintaining one's independence in activities of daily living, upper extremity impairments remain one of the most prevalent post-stroke deficits. To enhance the upper extremity motor recovery and performance among stroke survivors, two training paradigms in the fields of robotics therapy involving modifying haptic feedback were proposed: the error-augmentation (EA) and error-reduction (ER) paradigms. There is a lack of consensus, however, as to which of the two paradigms yields superior training effects. This systematic review aimed to determine (i) whether EA is more effective than conventional repetitive practice; (ii) whether ER is more effective than conventional repetitive practice and; (iii) whether EA is more effective than ER in improving post-stroke upper extremity motor recovery and performance. The study search and selection process as well as the ratings of methodological quality of the articles were conducted by two authors separately, and the results were then compared and discussed among the two reviewers. Findings were analyzed and synthesized using the level of evidence. By August 1st 2017, 269 articles were found after searching 6 databases, and 13 were selected based on criteria such as sample size, type of participants recruited, type of interventions used, etc. Results suggest, with a moderate level of evidence, that EA is overall more effective than conventional repetitive practice (motor recovery and performance) and ER (motor performance only), while ER appears to be no more effective than conventional repetitive practice. However, intervention effects as measured using clinical outcomes were under most instance not 'clinically meaningful' and effect sizes were modest. While stronger evidence is required to further support the efficacy of error modification therapies, the influence of factors related to the delivery of the intervention (such as intensity, duration) and personal factors (such as stroke severity and time of stroke onset) deserves further investigations as well.	Le Yu Liu, Youlin Li & Anouk Lamontagne	Journal of NeuroEngineering and Rehabilitation			2018	Si

143	Springer Link	Ultrasound of sacroiliac joints in spondyloarthritis: a systematic review	Ultrasound (US) is an accessible imaging technique with a possible role to diagnose active sacroiliitis, so this technique is projected as a promising diagnostic tool for the diagnosis of SpA. We analyse the available evidence about the use of US as a diagnostic tool in sacroiliitis in patients with SpA, by a systemic review of the literature fulfilling OMERACT criteria. A systematic literature search for original articles was carried out using four databases (Medline, Embase, Scopus and Web of Science). Data from studies were included only if participants had SpA and a US examination of sacroiliac joint (SIJ) was performed. The methodological quality of the studies was assessed using QUADAS-2 tool. Thirteen studies were included. All studies were observational, prospective and cross-sectional. In most articles (76.9%), the main US finding compatible with sacroiliitis evaluated was the presence of vascularisation (Doppler signals) with measurements of the resistive index (RI). The sensitivity and specificity analysis were performed in seven studies (58.8%) and were good, with a median of 90 and 89.2%, respectively. The studies showed a positive to moderate a strong correlation between the US and the gold standard but this was optimal only in four studies. In general, the agreement was good in all studies (≥ 0.80). The methods of evaluation of sacroiliitis vary between the studies included. To date, there is not enough evidence to support the use of ultrasound as a diagnostic method for sacroiliitis but it has potential to identify structural lesions at SIJ's level.	Marwin Gutierrez, Sheila Rodriguez, Carina Soto-Fajardo, Pedro Santos-Moreno, Hugo Sandoval, Chiara Bertolazzi & Carlos Pineda	Rheumatology International			2018	Si
144	Springer Link	Clinical Prediction Models	In this final chapter, we summarize some lessons learnt on development, validation, and updating of prediction models, based on the empirical experience from case studies as described in this book, and modeling experience in other medical prediction problems. We consider the essential elements to successful modeling: appropriate methods; sufficient sample size; emphasis on validation; using, not ignoring, subject matter knowledge. We also reflect further on modern machine learning techniques. Reporting guidelines and risk of bias tools are discussed. We end this chapter with a description of the case studies used throughout this book, where data sets are available through the book's website	Ewout W. Steyerberg				2019	Si
145	Springer Link	Prognostic models of diabetic microvascular complications: a systematic review and meta-analysis	Many prognostic models of diabetic microvascular complications have been developed, but their performances still varies. Therefore, we conducted a systematic review and meta-analysis to summarise the performances of the existing models.	Sigit Ari Saputro, Oraluck Pattanaprateep, Anuchate Pattanateepapon, Swekshya Karmacharya & Ammarin Thakkinsti	Systematic Reviews			2021	Si

				an					
146	Springer Link	Wearable Inertial Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review	Analysis of lower limb exercises is traditionally completed with four distinct methods: (1) 3D motion capture; (2) depth-camera-based systems; (3) visual analysis from a qualified exercise professional; and (4) self-assessment. Each method is associated with a number of limitations.	Martin O'Reilly, Brian Caulfield, Tomas Ward, William Johnston & Cailbhe Doherty	Sports Medicine			2018	Si
147	Springer Link	Factors associated with brain ageing - a systematic review	Brain age is a biomarker that predicts chronological age using neuroimaging features. Deviations of this predicted age from chronological age is considered a sign of age-related brain changes, or commonly referred to as brain ageing. The aim of this systematic review is to identify and synthesize the evidence for an association between lifestyle, health factors and diseases in adult populations, with brain ageing.	Jo Wrigglesworth, Phillip Ward, Ian H. Harding, Dinuli Nilaweera, Zimu Wu, Robyn L. Woods & Joanne Ryan	BMC Neurology			2021	Si
148	Springer Link	Prognostic models for knee osteoarthritis: a protocol for systematic review, critical appraisal, and meta-analysis	Osteoarthritis is the most common degenerative joint disease. It is associated with significant socioeconomic burden and poor quality of life, mainly due to knee osteoarthritis (KOA), and related total knee arthroplasty (TKA). Since early detection method and disease-modifying drug is lacking, the key of KOA treatment is shifting to disease prevention and progression slowing. The prognostic prediction models are called for to guide clinical decision-making. The aim of our review is to identify and characterize reported multivariable prognostic models for KOA about three clinical concerns: (1) the risk of developing KOA in the general population, (2) the risk of receiving TKA in KOA patients, and (3) the outcome of TKA in KOA patients who plan to receive TKA.	Jingyu Zhong, Liping Si, Guangcheng Zhang, Jiayu Huo, Yue Xing, Yangfan Hu, Huan Zhang & Weiwu Yao	Systematic Reviews			2021	Si
149	Springer Link	A question of trust: can we build an evidence base to gain trust in systematic review automation technologies?	Although many aspects of systematic reviews use computational tools, systematic reviewers have been reluctant to adopt machine learning tools.	Annette M. O'Connor, Guy Tsafnat, James Thomas, Paul Glasziou, Stephen B. Gilbert	Systematic Reviews			2019	Si

				& Brian Hutton					
150	Springer Link	Prediction modelling in the early detection of neonatal sepsis	Prediction modelling can greatly assist the health-care professionals in the management of diseases, thus sparking interest in neonatal sepsis diagnosis. The main objective of the study was to provide a complete picture of performance of prediction models for early detection of neonatal sepsis.	Puspita Sahu, Elstin Anbu Raj Stanly, Leslie Edward Simon Lewis, Krishnananda Prabhu, Mahadev Rao & Vijayanarayana Kunhikatta	World Journal of Pediatrics			2022	Si
151	Springer Link	Balance and proprioception impairment, assessment tools, and rehabilitation training in patients with total hip arthroplasty: a systematic review	Osteoarthritis and subsequent total hip arthroplasty (THA) lead to damages to hip joint mechanoceptors, which in turns lead to impairments in proprioception. One of the abilities mainly affected by an altered joint proprioception is balance. The aim of this work was to investigate the balance and proprioception impairments, current assessment tools, and rehabilitation training after THA.	Luciana Labanca, Francesca Ciardulli, Fabio Bonsanto, Nadia Sommella, Alberto Di Martino & Maria Grazia Benedetti	BMC Musculoskeletal Disorders			2021	Si
152	Springer Link	Integration of ultrasound in medical School: Effects on Physical Examination Skills of Undergraduates	Ultrasound (US) imaging has rapidly increased its application in almost every medical field. Many universities worldwide provide teaching of US for undergraduates in their curricula. Emerging evidence is supporting the use of ultrasonography to improve also non-US skills and knowledge of medical students.	Vittorio Oteri, Federica Occhipinti, Giorgia Gribaudo, Francesco Marastoni & Emanuele Chisari	Medical Science Educator			2020	Si
153	Springer Link	Efficacy of antiviral therapies for COVID-19: a systematic review of randomized controlled trials	Coronavirus disease 2019 (COVID-19) continues to pose a significant threat to public health worldwide. The purpose of this study was to review current evidence obtained from randomized clinical trials on the efficacy of antivirals for COVID-19 treatment.	Charan Thej Reddy Vegivinti, Kirk W.	BMC Infectious Diseases			2022	Si

				Evanson, Hannah Lyons, Izzet Akosman, Averi Barrett, Nicole Hardy					
154	Springer Link	The association between wearable activity monitor metrics and performance status in oncology: a systematic review	The expanding armamentarium of wearable activity monitors (WAMs) offers new opportunities to supplement physician-assessed performance status (PS) with real-life patient activity data. These data could guide clinical decision making or serve as a measure of treatment outcome. However, information on the association between physical activity (PA) and sedentary behavior (SB) monitored with wearables (i.e., WAM metrics) and PS in patients with cancer is needed. Therefore, we conducted a systematic review to examine the association between WAM metrics and PS in patients with cancer.	Milan Kos, Esther N. Pijnappel, Laurien M. Buffart, Britt R. Balvers, Caroline S. Kampshoff, Johanna W. Wilmink, Hanneke W. M. van Laarhoven & Martijn G. H. van Oijen	Supportive Care in Cancer			2021	Si
155	Springer Link	Use of handheld computers in clinical practice: a systematic review	Many healthcare professionals use smartphones and tablets to inform patient care. Contemporary research suggests that handheld computers may support aspects of clinical diagnosis and management. This systematic review was designed to synthesise high quality evidence to answer the question; Does healthcare professionals' use of handheld computers improve their access to information and support clinical decision making at the point of care?	Sharon Mickan, Helen Atherton, Nia Wyn Roberts, Carl Heneghan & Julie K Tilson	BMC Medical Informatics and Decision Making			2014	Si
156	Springer Link	Abstract analysis method facilitates filtering low-methodological quality and high-bias risk systematic reviews on psoriasis interventions	Article summaries' information and structure may influence researchers/clinicians' decisions to conduct deeper full-text analyses. Specifically, abstracts of systematic reviews (SRs) and meta-analyses (MA) should provide structured summaries for quick assessment. This study explored a method for determining the methodological quality and bias risk of full-text reviews using abstract information alone.	Francisco Gómez-García, Juan Ruano, Macarena Aguilar-Luque, Patricia Alcalde-	BMC Medical Research Methodology			2017	Si

				Mellado, Jesús Gay-Mimbrera					
157	Springer Link	Information Retrieval: A Biomedical and Health Perspective	This chapter provides an overview of biomedical and health information. It begins by defining the term information. This is followed by a discussion of theoretical aspects of information, followed by properties and classification of information. Next, the generation of scientific information is described, including primary and secondary scientific literature. That is followed by a discussion of electronic publishing. Finally, aspects of the use of biomedical and health information is described.	William Hersh				2020	Si
158	Springer Link	Stroke aetiological classification reliability and effect on trial sample size: systematic review, meta-analysis and statistical modelling	Inter-observer variability in stroke aetiological classification may have an effect on trial power and estimation of treatment effect. We modelled the effect of misclassification on required sample size in a hypothetical cardioembolic (CE) stroke trial.	Azmil H. Abdul-Rahim, David Alexander Dickie, Johann R. Selvarajah, Kennedy R. Lees & Terence J. Quinn On behalf of VISTA Collaborators	Trials			2019	Si
159	Springer Link	Current state of the art in the use of augmented reality in dentistry: a systematic review of the literature	The aim of the present systematic review was to screen the literature and to describe current applications of augmented reality.	Marco Farronato, Cinzia Maspero, Valentina Lanteri, Andrea Fama, Francesco Ferrati, Alessandro Pettenuzzo & Davide Farronato	BMC Oral Health			2019	Si
160	Springer Link	Systematic review automation technologies	Systematic reviews, a cornerstone of evidence-based medicine, are not produced quickly enough to support clinical practice. The cost of production, availability of the requisite expertise and timeliness are often quoted as major contributors for the delay. This detailed survey	Guy Tsafnat, Paul Glasziou,	Systematic Reviews			2014	Si

			of the state of the art of information systems designed to support or automate individual tasks in the systematic review, and in particular systematic reviews of randomized controlled clinical trials, reveals trends that see the convergence of several parallel research projects.	Miew Keen Choong, Adam Dunn, Filippo Galgani & Enrico Coiera					
161	Springer Link	Psychotherapy or medication for depression? Using individual symptom meta-analyses to derive a Symptom-Oriented Therapy (SOrT) metric for a personalised psychiatry	Antidepressant medication (ADM) and psychotherapy are effective treatments for major depressive disorder (MDD). It is unclear, however, if treatments differ in their effectiveness at the symptom level and whether symptom information can be utilised to inform treatment allocation. The present study synthesises comparative effectiveness information from randomised controlled trials (RCTs) of ADM versus psychotherapy for MDD at the symptom level and develops and tests the Symptom-Oriented Therapy (SOrT) metric for precision treatment allocation.	Nils Kappelmann, Martin Rein, Julia Fietz, Helen S. Mayberg, W. Edward Craighead, Boadie W. Dunlop	BMC Medicine			2020	Si
162	Springer Link	Machine learning for screening prioritization in systematic reviews: comparative performance of Abstrackr and EPPI-Reviewer	Improving the speed of systematic review (SR) development is key to supporting evidence-based medicine. Machine learning tools which semi-automate citation screening might improve efficiency. Few studies have assessed use of screening prioritization functionality or compared two tools head to head. In this project, we compared performance of two machine-learning tools for potential use in citation screening.	Amy Y. Tsou, Jonathan R. Treadwell, Eileen Erinoff & Karen Schoelles	Systematic Reviews			2020	Si
163	Springer Link	Application of artificial intelligence in diagnosis of osteoporosis using medical images: a systematic review and meta-analysis	Artificial intelligence (AI) is a potentially reliable assistant in the diagnosis of osteoporosis. This meta-analysis aims to assess the diagnostic accuracy of the AI-based systems using medical images. We searched PubMed and Web of Science from inception to June 15, 2020, for eligible articles that applied AI approaches to diagnosing osteoporosis using medical images. Quality and bias of the included studies were evaluated with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. The main outcome was the sensitivity and specificity of the performance of the AI-based systems. The data analysis utilized the R Foundation packages of “meta” for univariate analysis and Stata for bivariate analysis. Random effects model was utilized. Seven studies with 3186 patients were included in the meta-analysis. The overall risk of bias of the included studies was assessed as low. The pooled sensitivity was 0.96 (95% CI 0.93–1.00), and the pooled specificity was 0.95 (95% CI 0.91–0.99). However, high heterogeneity was found in this meta-analysis. The results supported that the AI-based systems had good accuracy in diagnosing osteoporosis. However, the high risk of	L. Gao, T. Jiao, Q. Feng & W. Wang	Osteoporosis International			2021	Si

			bias in patient selection and high heterogeneity in the meta-analysis made the conclusion less convincing. The application of AI-based systems in osteoporosis diagnosis needs to be further confirmed by more prospective studies in multi-centers including more random samples from complete patient types.						
164	Springer Link	Simulation for skills training in neurosurgery: a systematic review, meta-analysis, and analysis of progressive scholarly acceptance	At a time of significant global unrest and uncertainty surrounding how the delivery of clinical training will unfold over the coming years, we offer a systematic review, meta-analysis, and bibliometric analysis of global studies showing the crucial role simulation will play in training. Our aim was to determine the types of simulators in use, their effectiveness in improving clinical skills, and whether we have reached a point of global acceptance. A PRISMA-guided global systematic review of the neurosurgical simulators available, a meta-analysis of their effectiveness, and an extended analysis of their progressive scholarly acceptance on studies meeting our inclusion criteria of simulation in neurosurgical education were performed. Improvement in procedural knowledge and technical skills was evaluated. Of the identified 7405 studies, 56 studies met the inclusion criteria, collectively reporting 50 simulator types ranging from cadaveric, low-fidelity, and part-task to virtual reality (VR) simulators. In all, 32 studies were included in the meta-analysis, including 7 randomised controlled trials. A random effects, ratio of means effects measure quantified statistically significant improvement in procedural knowledge by 50.2% (ES 0.502; CI 0.355; 0.649, $p < 0.001$), technical skill including accuracy by 32.5% (ES 0.325; CI − 0.482; − 0.167, $p < 0.001$), and speed by 25% (ES − 0.25, CI − 0.399; − 0.107, $p < 0.001$). The initial number of VR studies (n = 91) was approximately double the number of refining studies (n = 45) indicating it is yet to reach progressive scholarly acceptance. There is strong evidence for a beneficial impact of adopting simulation in the improvement of procedural knowledge and technical skill. We show a growing trend towards the adoption of neurosurgical simulators, although we have not fully gained progressive scholarly acceptance for VR-based simulation technologies in neurosurgical education.	Joseph Davids, Susruta Manivannan, Ara Darzi, Stamatia Giannarou, Hutan Ashrafian & Hani J Marcus	Neurosurgical Review			2021	Si
165	Springer Link	Is there an association between non-alcoholic fatty liver disease and cognitive function? A systematic review	Non-alcoholic fatty liver disease (NAFLD) is represented as the most common liver disease worldwide. NAFLD is associated with metabolic risk factors underpinned by insulin resistance, inflammation and endothelial dysfunction, leading to extrahepatic changes in central nervous diseases such as cognitive impairment, Alzheimer's disease and dementia. The aim of the review is to explore the association between NAFLD and cognitive function.	Elena S. George, Surbhi Sood, Robin M. Daly & Sze-Yen Tan	BMC Geriatrics			2022	Si
166	Springer Link	Incorporating repeated	The incorporation of repeated measurements into multivariable prediction research may greatly enhance predictive performance.	Joost D. J. Plate,	BMC Medical Research			2019	Si

		measurements into prediction models in the critical care setting: a framework, systematic review and meta-analysis	However, the methodological possibilities vary widely and a structured overview of the possible and utilized approaches lacks. Therefore, we [1] propose a structured framework for these approaches, [2] determine what methods are currently used to incorporate repeated measurements in prediction research in the critical care setting and, where possible, [3] assess the added discriminative value of incorporating repeated measurements.	Rutger R. van de Leur, Luke P. H. Leenen, Falco Hietbrink, Linda M. Peelen & M. J. C. Eijkemans	Methodology				
167	Springer Link	Effectiveness of eccentric-biased exercise interventions in reducing the incidence of falls and improving functional performance in older adults: a systematic review	Exercise has been known to preserve and enhance functional performance in older adults. Eccentric exercise involves muscle contractions characterised by unique features such as lengthening of the muscle–tendon complex by a greater opposing force.	Durga Kulkarni, Sarah Gregory & Michelle Evans	European Geriatric Medicine			2021	Si
168	Springer Link	Mesenchymal stromal cells as a therapeutic intervention for COVID-19: a living systematic review and meta-analysis protocol	Mesenchymal stromal cells (MSCs) have significant immunomodulatory and tissue repair capabilities, mediated partly by conditioned media or through secreted extracellular vesicles (MSC-EVs). Infection with SARS-CoV-2 can cause mild to life-threatening illness due to activated immune responses that may be dampened by MSCs or their secretome. Many clinical studies of MSCs have been launched since the beginning of the global pandemic, however, few have been completed and most lack power to assess efficacy. Repeated systematic searches and meta-analyses are needed to understand, in real time, the extent of potential benefit in different patient populations as the evidence emerges.	Aidan M. Kirkham, Madeline Monaghan, Adrian J. M. Bailey, Risa Shorr, Manoj M. Lalu, Dean A. Fergusson & David S. Allan	Systematic Reviews			2021	Si
169	Springer Link	The Effect of Design Features on Relationship Quality with Embodied Conversational Agents: A Systematic Review	Embodied conversational agents (ECAs) are increasingly used in healthcare and other settings to improve self-management and provide companionship. Their ability to form close relationships with people is important for enhancing effectiveness and engagement. Several studies have looked at enhancing relationships with ECAs through design features focused on behaviours, appearance, or language. However, this evidence is yet to be systematically synthesized. This systematic review evaluates the effect of different design features on relationship quality with ECAs. A systematic search was conducted on electronic databases EMBASE, PsychInfo, PubMed, MEDLINE, Cochrane Library, SCOPUS, and Web of Science in January–February 2019. 43 studies were included for review that evaluated the effect of a design feature on relationship quality and social perceptions or behaviours towards an ECA.	Kate Loveys, Gabrielle Sebaratnam, Mark Sagar & Elizabeth Broadbent	International Journal of Social Robotics			2020	Si

			Results synthesize effective design features and lay a scientific framework for improving relationships with ECAs in healthcare and other applications. Risk of bias for included studies was generally low, however there were some limitations in the research quality pertaining to outcome measurement and the reporting of statistics. Further research is needed to understand how to make ECAs effective and engaging for all consumers.						
170	Springer Link	Efficacy of aerobic exercise on the cardiometabolic and renal outcomes in patients with chronic kidney disease: a systematic review of randomized controlled trials	Several randomized controlled trials (RCTs) have demonstrated the cardiometabolic effects of aerobic exercise in the general population and in patients with cardiovascular diseases. However, the efficacy of aerobic exercise in patients with chronic kidney disease (CKD) remains to be elucidated.	Ryohei Yamamoto, Takafumi Ito, Yasuyuki Nagasawa, Kosuke Matsui, Masahiro Egawa, Masayoshi Nanami, Yoshitaka Isaka & Hirokazu Okada	Journal of Nephrology			2021	Si
171	Springer Link	Eccentric Strength Assessment of Hamstring Muscles with New Technologies: a Systematic Review of Current Methods and Clinical Implications	Given the severe economic and performance implications of hamstring injuries, there are different attempts to identify their risk factors for subsequently developing injury prevention strategies to reduce the risk of these injuries. One of the strategies reported in the scientific literature is the application of interventions with eccentric exercises. To verify the effectiveness of these interventions, different eccentric strength measurements have been used with low-cost devices as alternatives to the widespread used isokinetic dynamometers and the technically limited handheld dynamometers. Therefore, the purpose of the present systematic review was to summarize the findings of the scientific literature related to the evaluation of eccentric strength of hamstring muscles with these new technologies.	João Gustavo Claudino, Carlos Alberto Cardoso Filho, Natália Franco Netto Bittencourt, Luiz Gilherme Gonçalves	Sports Medicine			2021	Si
172	Springer Link	The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults	Step-count monitors (pedometers, body-worn trackers and smartphone applications) can increase walking, helping to tackle physical inactivity. We aimed to assess the effect of step-count monitors on physical activity (PA) in randomised controlled trials (RCTs) amongst community-dwelling adults; including longer-term effects, differences between step-count monitors, and between intervention components.	Umar A. R. Chaudhry, Charlotte Wahlich, Rebecca Fortescue, Derek G. Cook, Rachel Knightly &	International Journal of Behavioral Nutrition and Physical Activity			2020	Si

				Tess Harris					
173	Springer Link	Risk of bias judgements and strength of conclusions in meta-evidence from the Cochrane Colorectal Cancer Group	The Cochrane Collaboration records risk of bias (ROB) judgements on the original studies it analyses. The aim of this review is to perform an audit of all literature produced by the Cochrane Colorectal Cancer Group (CCCG), focusing on whether intervention type has any relationship with ROB and the ability of a review to inform clinical practice.	John Delaney, Rebecca Cui & Alexander Engel	Systematic Reviews			2019	Si
174	Springer Link	How to conduct systematic reviews more expeditiously?	Healthcare consumers, researchers, patients and policy makers increasingly use systematic reviews (SRs) to aid their decision-making process. However, the conduct of SRs can be a time-consuming and resource-intensive task. Often, clinical practice guideline developers or other decision-makers need to make informed decisions in a timely fashion (e.g. outbreaks of infection, hospital-based health technology assessments). Possible approaches to address the issue of timeliness in the production of SRs are to (a) implement process parallelisation, (b) adapt and apply innovative technologies, and/or (c) modify SR processes (e.g. study eligibility criteria, search sources, data extraction or quality assessment). Highly parallelised systematic reviewing requires substantial resources to support a team of experienced information specialists, reviewers and methodologists working alongside with clinical content experts to minimise the time for completing individual review steps while maximising the parallel progression of multiple steps. Effective coordination and management within the team and across external stakeholders are essential elements of this process. Emerging innovative technologies have a great potential for reducing workload and improving efficiency of SR production. The most promising areas of application would be to allow automation of specific SR tasks, in particular if these tasks are time consuming and resource intensive (e.g. language translation, study selection, data extraction). Modification of SR processes involves restricting, truncating and/or bypassing one or more SR steps, which may risk introducing bias to the review findings. Although the growing experiences in producing various types of rapid reviews (RR) and the accumulation of empirical studies exploring potential bias associated with specific SR tasks have contributed to the methodological development for expediting SR production, there is still a dearth of research examining the actual impact of methodological modifications and comparing the findings between RRs and SRs. This evidence would help to inform as to which SR tasks can be accelerated or truncated and to what degree, while maintaining the validity of review findings. Timely delivered SRs can be of value in informing healthcare decisions and recommendations, especially when there is practical urgency and there is no other relevant synthesised evidence.	Alexander Tsertsvadze, Yen-Fu Chen, David Moher, Paul Sutcliffe & Noel McCarthy	Systematic Reviews			2015	Si

175	Springer Link	Effectiveness of health-oriented leadership interventions for improving health and wellbeing of employees: a systematic review	To identify and summarize the evidence for the effect of health-oriented leadership interventions on health and well-being outcomes at the employee level following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (Moher et al. 2009).	Indra Dannheim, Helena Ludwig-Walz, Anette E. Buyken, Valerie Grimm & Anja Kroke	Journal of Public Health			2021	Si
176	Springer Link	ICT technologies as new promising tools for the managing of frailty: a systematic review	Frailty is a major health issue as it encompasses functional decline, physical dependence, and increased mortality risk. Recent studies explored Information and Communication Technology (ICT) interventions as alternatives to manage frailty in older persons. The aim of the present systematic review was to synthesize current evidence on ICT application within the complex models of frailty care in older people.	Alessia Gallucci, Pietro Davide Trimarchi, Carlo Abbate, Cosimo Tuena, Elisa Pedroli, Fabrizia Lattanzio, Marco Stramba-Badiale, Matteo Cesari & Fabrizio Giunco	Aging Clinical and Experimental Research			2021	Si
177	Springer Link	Is serum C-terminal telopeptide cross-link of type 1 collagen a reliable parameter for predicting the risk of medication-related osteonecrosis of the jaws? A systematic review and meta-analysis of diagnostic test accuracy	To determine the usefulness of Serum C-terminal telopeptide cross-link of type 1 collagen (sCTX) as a preoperative marker for predicting the risk of developing medication-related osteonecrosis of the jaws (MRONJ) after invasive oral surgery in patients on antiresorptive medication.	Bassel Traboulsi-Garet, Adrià Jorba-García, Octavi Camps-Font, Fabio Abreu Alves, Rui Figueiredo & Eduard Valmaseda-Castellón	Clinical Oral Investigations			2022	Si

178	Springer Link	The applications of DNA methylation as a biomarker in kidney transplantation: a systematic review	Although kidney transplantation improves patient survival and quality of life, long-term results are hampered by both immune- and non-immune-mediated complications. Current biomarkers of post-transplant complications, such as allograft rejection, chronic renal allograft dysfunction, and cutaneous squamous cell carcinoma, have a suboptimal predictive value. DNA methylation is an epigenetic modification that directly affects gene expression and plays an important role in processes such as ischemia/reperfusion injury, fibrosis, and alloreactive immune response. Novel techniques can quickly assess the DNA methylation status of multiple loci in different cell types, allowing a deep and interesting study of cells' activity and function. Therefore, DNA methylation has the potential to become an important biomarker for prediction and monitoring in kidney transplantation.	Iacopo Cristoferi, Tommaso Antonio Giacon, Karin Boer, Myrthe van Baardwijk, Flavia Neri, Manuela Campisi, Hendrikus J. A. N. Kimenai, Marian C. Clahsen - van Groningen, Sofia Pavanello, Lucrezia Furian & Robert C. Minnee	Clinical Epigenetics			2022	Si
179	Springer Link	Neuroimaging and analytical methods for studying the pathways from mild cognitive impairment to Alzheimer's disease: protocol for a rapid systematic review	Alzheimer's disease (AD) is a neurodegenerative disorder commonly associated with deficits of cognition and changes in behavior. Mild cognitive impairment (MCI) is the prodromal stage of AD that is defined by slight cognitive decline. Not all with MCI progress to AD dementia. Thus, the accurate prediction of progression to Alzheimer's, particularly in the stage of MCI could potentially offer developing treatments to delay or prevent the transition process. The objective of the present study is to investigate the most recent neuroimaging procedures in the domain of prediction of transition from MCI to AD dementia for clinical applications and to systematically discuss the machine learning techniques used for the prediction of MCI conversion.	Maryam Ahmadzadeh, Gregory J. Christie, Theodore D. Cosco & Sylvain Moreno	Systematic Reviews			2020	Si
180	Springer Link	Efficacy of resistance training during adjuvant chemotherapy and radiation therapy in cancer care: a systematic review and meta-analysis	To determine the effect of resistance training during adjuvant chemotherapy and radiation therapy in cancer patients on measures of lean mass and muscle strength. Secondary aims were to analyse the prescription and tolerability of supervised resistance training in this population.	Aoife McGovern, Nicholas Mahony, David Mockler & Neil Fleming	Supportive Care in Cancer			2022	Si

181	Springer Link	Patient centred variables with univariate associations with unplanned ICU admission: a systematic review	Multiple predictive scores using Electronic Patient Record data have been developed for hospitalised patients at risk of clinical deterioration. Methods used to select patient centred variables for inclusion in these scores varies. We performed a systematic review to describe univariate associations with unplanned Intensive Care Unit (ICU) admission with the aim of assisting model development for future scores that predict clinical deterioration.	James Malycha, Timothy Bonnici, David A. Clifton, Guy Ludbrook, J. Duncan Young & Peter J. Watkinson	BMC Medical Informatics and Decision Making			2019	Si
182	Springer Link	Common Elements of Practice, Process and Implementation in Out-of-School-Time Academic Interventions for At-risk Children: a Systematic Review	Academic achievement is a strong preventive factor against marginalization. Children at risk of academic failure and drop out can benefit from out-of-school-time academic (OSTA) interventions. Wide-scaled implementation and sustainment of effective interventions remain a struggle across education, welfare, and health. The need for approaches to increase implementability, effectiveness, and efficiency of interventions is pressing. Advancements in the field of education and mental health suggest identifying and studying discrete elements that are common across interventions for the purpose of hypothesis generation, intervention optimization, design improvement, and implementation. This review identified OSTA interventions for primary school children at risk of academic failure. Common elements methodology was used to code practice elements (n = 62), process elements (n = 49), and implementation elements (n = 36) in 30 effective and six ineffective OSTA interventions in matrices. Based on frequency counts, common practice, process, and implementation elements across the interventions were identified, and given frequency count values (FV) reflecting how often elements were included in effective studies as opposed to in ineffective studies. The five common practice elements with the highest FVs were homework support, training in positive parental school involvement, positive reinforcement, structured tutoring, and psychoeducation. The most common process element was regular support to intervention receiver, and the most common implementation element was quality monitoring. Common combinations of elements were also identified and given FVs. Results from this review can inform efforts to design or optimize OSTA interventions, and inform education, implementation, and practice to improve academic achievement for children at risk.	Thomas Engell, Benedicte Kirkøen, Karianne Thune Hammerstrøm, Hege Kornør, Kristine Horseng Ludvigsen & Kristine Amlund Hagen	Prevention Science			2020	Si
183	Springer Link	Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review	High levels of sedentary behaviour (SB) are associated with negative health consequences. Technology enhanced solutions such as mobile applications, activity monitors, prompting software, texts, emails and websites are being harnessed to reduce SB. The aim of this paper is to evaluate the effectiveness of such technology enhanced interventions aimed at reducing SB in healthy adults and to examine the behaviour change techniques (BCTs) used.	Aoife Stephenson, Suzanne M. McDonough, Marie	International Journal of Behavioral Nutrition and Physical Activity			2017	Si

		and meta-analysis		H. Murphy, Chris D. Nugent & Jacqueline L. Mair					
184	Springer Link	Evidence synthesis of types and intensity of therapeutic land-based exercises to reduce pain in individuals with knee osteoarthritis	The objective of this study is to construct an evidence synthesis to identify the types of land-based exercises most investigated in the current literature, the intervention duration, frequency of the programs and the exercises which are most frequently implemented. A search was performed on the reference list of included and excluded studies of one systematic review, on land-based exercises for knee osteoarthritis and, an updated search of The Cochrane Library, Embase, CINAHL and PEDro was completed. Two authors independently selected the studies and a third author was consulted for an additional opinion. The inclusion criteria were male or female with tibiofemoral knee osteoarthritis, land-based exercises, non-exercise control group and randomized clinical trials. The exclusion criteria were mixed diagnosis or comparison to other types of exercise. The data were extracted by two authors. Fifty-five full-text articles were included. Strengthening, proprioception and aerobic exercises resulted in significant pain reduction. The intervention durations which were significant for pain reduction were either the period of 8–11 weeks or 12–15 weeks. The frequency of three times per week was found significant in comparison to a non-exercise control group. The results, which formed an evidence synthesis, demonstrate that there is substantial evidence regarding the benefits of strengthening exercises to reduce pain in knee osteoarthritis patients. Based on the included studies analysis, exercises should be performed three times weekly for a duration of 8–11 or 12–15 weeks. Health professionals working with knee osteoarthritis patients can use this evidence synthesis as a fast and pragmatic instrument to obtain information about several effective types of exercises for pain reduction.	Aline Mizusaki Imoto, Jordi Pardo Pardo, Lucie Brosseau, Jade Taki, Brigit Desjardins, Odette Thevenot, Eduardo Franco & Stella Peccin	Rheumatology International			2019	Si
185	Springer Link	Effectiveness of web-based feedback interventions for people with overweight and obesity: systematic review and network meta-analysis of randomized controlled trials	Web-based delivered interventions have become an innovative option to treat health problems, like obesity. The aim of this systematic review and network meta-analysis was to analyze the effectiveness of web-based behavioral treatments for adults with overweight and obesity. Web-based interventions and comparison interventions (traditional weight control programs) were classified according to the following feedback characteristics: frequency, personalization, and provider (human versus machine).	Carmen Varela, Camila Oda-Montecinos, Ana Andrés & Carmina Saldaña	Journal of Eating Disorders			2021	Si
186	Springer Link	Helicopter emergency medical services use of thoracic point of care ultrasound for	Auscultating for breath sounds to assess for pneumothorax in the helicopter emergency medical services (HEMS) settings can be extremely challenging. Thoracic point of care ultrasound (POCUS) offers a seemingly more useful visual (rather than audible)	Edward Griffiths	Scandinavian Journal of Trauma, Resuscitation			2021	Si

		pneumothorax: a systematic review and meta-analysis	alternative. This review critically and quantitatively evaluates the use of thoracic POCUS for pneumothorax in the HEMS setting.		and Emergency Medicine				
187	Springer Link	Artificial intelligence (AI) real-time detection vs. routine colonoscopy for colorectal neoplasia: a meta-analysis and trial sequential analysis	Studies analyzing artificial intelligence (AI) in colonoscopies have reported improvements in detecting colorectal cancer (CRC) lesions, however its utility in the realworld remains limited. In this systematic review and meta-analysis, we evaluate the efficacy of AI-assisted colonoscopies against routine colonoscopy (RC).	Smit S. Deliwala, Kewan Hamid, Mahmoud Barbarawi, Harini Lakshman, Yazan Zayed, Pujan Kandel, Srikanth Malladi, Adiraj Singh, Ghassan Bachuwa, Grigoriy E. Gurvits & Saurabh Chawla	International Journal of Colorectal Disease			2021	Si
188	Springer Link	Delirium and other neuropsychiatric manifestations of COVID-19 infection in people with preexisting psychiatric disorders: a systematic review	Psychiatric disorders increase risk of neuropsychiatric disease and poor outcomes, yet little is known about the neuropsychiatric manifestations of COVID-19 in the psychiatric population. The primary objective is to synthesize neuropsychiatric outcomes of COVID-19 in people with preexisting psychiatric disorders.	Emma A. van Reekum, Tea Rosic, Anjali Sergeant, Nitika Sanger, Myanca Rodrigues, Reid Rebinsky, Balpreet Panesar, Eve Deck, Nayeon Kim, Julia Woo, Alessia D'Elia, Alannah Hillmer, Alexander	Journal of Medical Case Reports			2021	Si

				Dufort, Stephanie Sanger, Lehana Thabane, Lawrence Mbuagbaw & Zainab Samaan					
189	Springer Link	How should long-term free-living physical activity be targeted after stroke? A systematic review and narrative synthesis	Increasing physical activity (PA) levels (regular movement such as walking and activities of daily living) and reducing time spent sedentary improves cardiovascular health and reduces morbidity and mortality. Fewer than 30% of independently mobile stroke survivors undertake recommended levels of PA. Sedentary behaviour is also high in this population. We aimed to systematically review the study characteristics and the promise of interventions targeting free-living PA and/or sedentary behaviour in adult stroke survivors.	Sarah A. Moore, Nina Hrisos, Darren Flynn, Linda Errington, Christopher Price & Leah Avery	International Journal of Behavioral Nutrition and Physical Activity			2018	Si
190	Springer Link	Efficacy of digital single-operator cholangioscopy in the visual interpretation of indeterminate biliary strictures: a systematic review and meta-analysis	Indeterminate biliary strictures remain a significant diagnostic challenge. Digital single-operator cholangioscopy (D-SOC) incorporates digital imaging which enables higher resolution for better visualization and diagnosis of biliary pathology. We aimed to conduct a systematic review and meta-analysis of available literature in an attempt to determine the efficacy of D-SOC in the visual interpretation of indeterminate biliary strictures.	Pedro Victor Aniz Gomes de Oliveira, Diogo Turiani Hourneaux de Moura, Igor Braga Ribeiro, Ahmad Najdat Bazarbashi, Tomazo Antonio Prince Franzini, Marcos Eduardo Lera dos Santos, Wanderley Marques Bernardo & Eduardo	Surgical Endoscopy			2020	Si

				Guimarães Hourneaux de Moura					
191	Springer Link	A focus on cross-purpose tools, automated recognition of study design in multiple disciplines, and evaluation of automation tools: a summary of significant discussions at the fourth meeting of the International Collaboration for Automation of Systematic Reviews (ICASR)	A focus on cross-purpose tools, automated recognition of study design in multiple disciplines, and evaluation of automation tools: a summary of significant discussions at the fourth meeting of the International Collaboration for Automation of Systematic Reviews (ICASR)	Annette M. O'Connor, Paul Glasziou, Michele Taylor, James Thomas, René Spijker & Mary S. Wolfe	Systematic Reviews			2020	Si
192	Springer Link	Virtual follow-up and care for patients with cardiac electronic implantable devices: protocol for a systematic review	Capacity to deliver outpatient care for patients with cardiac implantable electronic devices (CIEDs) may soon be outweighed by need. This systematic review aims to investigate the comparative effectiveness, safety, and cost for virtual or remote clinic interventions for patients with CIEDs and explores how outcomes may be influenced by patient or system factors in-depth.	Shannon E. Kelly, Tammy J. Clifford, Doug Coyle, Janet Martin, Vivian Welch, Becky Skidmore, David Birnie, Ratika Parkash, Anthony S. L. Tang & George A. Wells	Systematic Reviews			2020	Si
193	Springer Link	Between-laboratory reproducibility of time-lapse embryo selection using qualitative and quantitative	To investigate the between-laboratory reproducibility of embryo selection/deselection effectiveness using qualitative and quantitative time-lapse parameters.	Yanhe Liu, Fang Qi, Phillip Matson, Dean E. Morbeck,	Journal of Assisted Reproduction and Genetics			2020	Si

		parameters: a systematic review and meta-analysis		Ben W. Mol, Sai Zhao & Masoud Afnan					
194	Springer Link	The Effect of Load and Volume Autoregulation on Muscular Strength and Hypertrophy: A Systematic Review and Meta-Analysis	Autoregulation has emerged as a potentially beneficial resistance training paradigm to individualize and optimize programming; however, compared to standardized prescription, the effects of autoregulated load and volume prescription on muscular strength and hypertrophy adaptations are unclear. Our objective was to compare the effect of autoregulated load prescription (repetitions in reserve-based rating of perceived exertion and velocity-based training) to standardized load prescription (percentage-based training) on chronic one-repetition maximum (1RM) strength and cross-sectional area (CSA) hypertrophy adaptations in resistance-trained individuals. We also aimed to investigate the effect of volume autoregulation with velocity loss thresholds ≤ 25% compared to > 25% on 1RM strength and CSA hypertrophy.	Landyn M. Hickmott, Philip D. Chilibeck, Keely A. Shaw & Scotty J. Butcher	Sports Medicine - Open			2022	Si
195	Springer Link	The PRISMA 2020 statement: an updated guideline for reporting systematic reviews		Matthew J. Page, Joanne E. McKenzie, …David Moher	Systematic Reviews			2021	Si
196	Springer Link	A Classification of Components of Workplace Disability Management Programs: Results from a Systematic Review	Purpose: This paper presents results from a Campbell systematic review on the nature and effectiveness of workplace disability management programs (WPDM) promoting return to work (RTW), as implemented and practiced by employers. A classification of WPDM program components, based on the review results, is proposed. Methods: Twelve databases were searched between 1948 to July 2010 for peer-reviewed studies of WPDM programs provided by employers to re-entering workers with occupational or non-occupational illnesses or injuries. Screening of articles, risk of bias assessment and data extraction were conducted in pairs of reviewers. Studies were clustered around various dimensions of the design and context of programs. Results: 16,932 records were identified by the initial search. 599 papers were assessed for relevance. Thirteen studies met inclusion criteria. Twelve peer reviewed articles (two non-randomized studies, and ten single group experimental before and after studies), including ten different WPDM programs informed the synthesis of results. Narrative descriptions of the included program characteristics provided insight on program scope, components, procedures and human resources involved. However, there were insufficient data on the characteristics of the sample and the effect sizes were uncertain. A taxonomy classifying policies and practices around WPDM programs is proposed. Conclusion: There is insufficient evidence to draw conclusions on the effectiveness of employer provided WPDM programs promoting	U. Gensby, M. Labriola, E. Irvin, B. C. Amick III & T. Lund	Journal of Occupational Rehabilitation			2014	Si

			RTW. It was not possible to determine if specific program components or specific sets of components are driving effectiveness. The proposed taxonomy may guide future WPDM program evaluation and clarify the setup of programs offered to identify gaps in existing company strategies.						
197	Springer Link	What do we know about the effects of exposure to 'Low alcohol' and equivalent product labelling on the amounts of alcohol, food and tobacco people select and consume? A systematic review	Explicit labelling of lower strength alcohol products could reduce alcohol consumption by attracting more people to buy and drink such products instead of higher strength ones. Alternatively, it may lead to more consumption due to a 'self-licensing' mechanism. Equivalent labelling of food or tobacco (for example "Low fat" or "Low tar") could influence consumption of those products by similar mechanisms. This systematic review examined the effects of 'Low alcohol' and equivalent labelling of alcohol, food and tobacco products on selection, consumption, and perceptions of products among adults.	Ian Shemilt, Vivien Hendry & Theresa M. Marteau	BMC Public Health			2017	Si
198	Springer Link	Interventions to prevent or treat childhood obesity in Māori & Pacific Islanders: a systematic review	Māori and Pacific Islander people are a priority population originating from Australasia. Māori and Pacific Islander children exhibit greater risk of obesity and associated morbidities compared to children of other descent, secondary to unique cultural practices and socioeconomic disadvantage. Despite these known risk factors, there is limited synthesised evidence for preventing and treating childhood obesity in this unique population. The objective of this systematic review was to identify and evaluate global prevention or treatment interventions for overweight or obesity that targeted Māori and Pacific Islander children and adolescents (aged 2–17 years).	Robyn Littlewood, Oliver J. Canfell & Jacqueline L. Walker	BMC Public Health			2020	Si
199	Springer Link	Self-management interventions for chronic kidney disease: a systematic review and meta-analysis	Self-management intervention aims to facilitate an individual's ability to make lifestyle changes. The effectiveness of this intervention in non-dialysis patients with chronic kidney disease (CKD) is limited. In this study, we applied a systematic review and meta-analysis to investigate whether self-management intervention improves renoprotection for non-dialysis chronic kidney disease.	Suyuan Peng, Jiawei He, Jiasheng Huang, Longwei Lun, Jiahao Zeng, Shan Zeng, La Zhang, Xusheng Liu & Yifan Wu	BMC Nephrology			2019	Si
200	Springer Link	Magnetic resonance imaging for forensic age estimation in living children and young adults: a systematic review	The use of MRI in forensic age estimation has been explored extensively during the last decade. The authors of this paper synthesized the available MRI data for forensic age estimation in living children and young adults to provide a comprehensive overview that can guide age estimation practice and future research. To do so, the authors searched MEDLINE, Embase and Web of Science, along with cited and citing articles and study registers. Two	Jannick De Tobel, Jeroen Bauwens, Griet I. L. Parmentier, Ademir	Pediatric Radiology			2020	Si

			authors independently selected articles, conducted data extraction, and assessed risk of bias. They considered study populations including living subjects up to 30 years old. Fifty-five studies were included in qualitative analysis and 33 in quantitative analysis. Most studies had biases including use of relatively small European (Caucasian) populations, varying MR approaches and varying staging techniques. Therefore, it was not appropriate to pool the age distribution data. The authors found that reproducibility of staging was remarkably lower in clavicles than in any other anatomical structure. Age estimation performance was in line with the gold standard, radiography, with mean absolute errors ranging from 0.85 years to 2.0 years. The proportion of correctly classified minors ranged from 65% to 91%. Multifactorial age estimation performed better than that based on a single anatomical site. The authors found that more multifactorial age estimation studies are necessary, together with studies testing whether the MRI data can safely be pooled. The current review results can guide future studies, help medical professionals to decide on the preferred approach for specific cases, and help judicial professionals to interpret the evidential value of age estimation results.	Franco, Nele S. Pauwels, Koenraad L. Verstraete & Patrick W. Thevissen					
201	Springer Link	Risk of Human Illness from Recreational Exposure to Microbial Pathogens in Freshwater Bodies: A Systematic Review	Recreational use of natural freshwater bodies poses a risk to human health, although the risks associated with different types of exposure in freshwater are not clear. A systematic review was carried out to identify the risks of adverse health outcomes among individuals exposed through primary contact recreation compared to minimal contact recreation in different types of freshwater bodies. Among 5858 potentially relevant titles and abstracts, 35 were eligible for final inclusion. We included studies that reported the type of freshwater exposure and health outcomes, were peer-reviewed, and had full-text articles available in English. Across all studies, general gastrointestinal illness was the most commonly reported (77.1%), followed by skin (37.1%) and ear-related infection (34.3%), respiratory illness (31.4%), and eye-related illness (25.7%). Most studies reported adverse health outcomes from primary contact recreation (85.7%) in comparison to secondary contact activities (17.1%) and no-contact recreation (2.9%). More than half of the included studies (54.3%) focussed on lakes. The continued focus on primary contact recreation in lakes is a concern, given the growing global trend in secondary and non-contact recreation associated with urban wetlands. A better understanding of risks associated with recreational contact with freshwater in these settings is needed.	Ripon Kumar Adhikary, Md Sarower-E. Mahfuj, Danswell Starrs, Barry Croke, Kathryn Glass & Aparna Lal	Exposure and Health			2021	Si
202	Springer Link	Clarifying differences between reviews within evidence ecosystems	This paper builds on a 2012 paper by the same authors which argued that the types and brands of systematic review do not sufficiently differentiate between the many dimensions of different review questions and review methods (Gough et al., Syst Rev 1:28, 2012). The current paper extends this argument by considering the dynamic contexts, or 'evidence ecosystems', within which reviews are undertaken; the fact that these ecosystems are constantly	David Gough, James Thomas & Sandy Oliver	Systematic Reviews			2019	Si

			changing; and the relevance of this broader context for understanding 'dimensions of difference' in the unfolding development and refinement of review methods. The concept of an evidence ecosystem is used to consider particular issues within the three key dimensions of difference outlined in the 2012 paper of (1) review aims and approach, (2) structure and components of reviews, and (3) breadth, depth, and 'work done' by reviews.						
203	Springer Link	Neuropsychological Interventions for Cancer-Related Cognitive Impairment: A Network Meta-Analysis of Randomized Controlled Trials	The aim of this network meta-analysis was to evaluate the comparative effects of neuropsychological interventions for cancer-related cognitive impairment (CRCI), and to rank the best intervention options for adult cancer patients with CRCI. Twenty-seven eligible randomized controlled trials (RCTs) were searched, and a total of six interventions identified: cognitive behavioral therapies (CBT), cognitive rehabilitation (CR), cognitive training (CT), meditation/mindfulness-based interventions, psychoeducation, and supportive care. In terms of effectiveness, the relative effect size of CBT, CR, and CT in managing subjective cognition had statistically significant differences – 0.94 (0.43–1.44), 0.54 (0.03–1.05), and 0.47 (0.13–0.81), respectively. The most effective interventions to manage the objective cognition of attention were meditation or mindfulness-based interventions: intervention effect size was 0.58 (0.24–0.91). The relative effect size of CT had a statistically significant difference in managing verbal memory, and the intervention effect size was 1.16 (0.12–2.20). The relative effect size of psychoeducation in managing executive function compared with control had a statistically significant difference, which was 0.56 (0.26–0.86). For managing information processing speed, the most effective intervention was CT and the effect size was -0.58 (-1.09—-0.06). This network meta-analysis found that CT is the most effective intervention for managing the objective cognition of verbal memory and processing speed; meditation/mindfulness-based interventions may be the best option for enhancing attention; psychoeducation is the most effective intervention for managing executive function; CT may be the best option for managing verbal fluency as the intervention ranking probability. For the management of subjective cognition, CBT may be the most effective intervention.	Andy S. K. Cheng, Xiaoming Wang, Niu Niu, Minyu Liang & Yingchun Zeng	Neuropsychology Review			2022	Si
204	Springer Link	Feasibility and acceptability of living systematic reviews: results from a mixed-methods evaluation	Living systematic reviews (LSRs) offer an approach to keeping high-quality evidence synthesis continually up to date, so the most recent, relevant and reliable evidence can be used to inform policy and practice, resulting in improved quality of care and patient health outcomes. However, they require modifications to authoring and editorial processes and pose technical and publishing challenges. Several teams within Cochrane and the international Living Evidence Network have been piloting living systematic reviews.	Tanya Millard, Anneliese Synnot, Julian Elliott, Sally Green, Steve McDonald	Systematic Reviews			2019	Si

				& Tari Turner					
205	Springer Link	Microcirculatory perfusion disturbances following cardiopulmonary bypass: a systematic review	Microcirculatory perfusion disturbances are associated with increased morbidity and mortality in patients undergoing cardiac surgery with cardiopulmonary bypass (CPB). Technological advancements made it possible to monitor sublingual microcirculatory perfusion over time. The goal of this review is to provide an overview of the course of alterations in sublingual microcirculatory perfusion following CPB. The secondary goal is to identify which parameter of sublingual microcirculatory perfusion is most profoundly affected by CPB.	Matthijs M. den Os, Charissa E. van den Brom, Anoek L. I. van Leeuwen & Nicole A. M. Dekker	Critical Care			2020	Si
206	Springer Link	Treatment of neurofibromatosis 1-associated malignant peripheral nerve sheath tumors: a systematic review	Malignant peripheral nerve sheath tumors (MPNST) are a rare and aggressive group of tumors that are challenging to treat. Neurofibromatosis type 1 (NF-1)-associated MPNSTs have been associated with poorer clinical outcomes. The treatment options for NF-1-associated MPNSTs broadly include surgery (SG), chemotherapy (CT), and adjuvant radiotherapy (RT). Overall, the role and efficacy of CT and RT are unclear. Examination of existing literature for studies reporting on NF-1-associated MPNSTs and respective treatment-related outcomes was conducted. We conducted a systematic review according to PRISMA guidelines in PubMed/Medline and Cochrane databases of studies which reported treatment-specific outcomes in NF-1-associated MPNSTs. The literature search found 444 records after removal of duplicates. The present study included 50 patients across 12 observational studies. All of the included studies reported data on overall survival (OS 52%, n = 26/50) but mean follow-up in months among the studies and among patients varied widely, between 10.85 (SD, ± 10.38) and 192 (SD, ± 98.22). From the included studies, patients underwent either SG alone (n = 21), SG + CT (n = 10), SG + RT (n = 7), or SG + CT + RT (n = 12). The quality of evidence in the literature regarding optimal treatment options for NF-1-associated MPNSTs remains tenuous. Future retrospective and prospective comparative trials should consider adherence to a set of reporting guidelines to improve the quality of evidence in the literature with respect to individual treatment-related outcomes. The need for prospective multi-institutional efforts cannot be overstated.	Muhibullah S. Tora, Dimitrios Xenos, Pavlos Texakalidis & Nicholas M. Boulis	Neurosurgical Review			2020	Si
207	Springer Link	Evaluating characteristics of PROSPERO records as predictors of eventual publication of non-Cochrane systematic reviews: a meta-epidemiological study protocol	Epidemiology and the reporting characteristics of systematic reviews (SRs) and meta-analyses (MAs) are well known. However, no study has analyzed the influence of protocol features on the probability that a study's results will be finally reported, thereby indirectly assessing the reporting bias of International Prospective Register of Systematic Reviews (PROSPERO) registration records.	Juan Ruano, Francisco Gómez-García, Jesús Gay-Mimbrera, Macarena	Systematic Reviews			2018	Si

				Aguilar-Luque, José Luis Fernández-Rueda, Jesús Fernández-Chaichio, Patricia Alcalde-Mellado, Pedro J. Carmona-Fernandez, Juan Luis Sanz-Cabanillas, Isabel Viguera-Guerra, Francisco Franco-García, Manuel Cárdenas-Aranzana, José Luis Hernández Romero, Marcelino Gonzalez-Padilla, Beatriz Isla-Tejera & Antonio Velez Garcia-Nieto					
208	Springer Link	Principles and Practice of Systematic Reviews and Meta-Analysis	Accurate data extraction and their synthesis form the basis of appropriate conclusions of a systematic review. Systematic reviewers should extract ALL data relevant to the review question, not just the outcome data. Data to be extracted include baseline characteristics of study participants, information related to study methodology and outcomes and other relevant information. If published articles have given the results using figures instead of actual numbers, specialised software that convert images to pixel values may be utilised to obtain the actual data values. Tools such as Plot Digitizer, WebPlotDigitizer, Engauge, Dexter, Ycasd and	Kwi MoonEmail authorShripada Rao				2021	Si

			GetData Graph Digitizer can be used for this purpose. When unable to extract data from available reports or to seek clarifications, the reviewers could contact the original investigators. Data extraction should be performed using pre-piloted forms independently by at least two reviewers to ensure accuracy. A high level of diligence is required to minimise errors during the stage of data extraction.						
209	Springer Link	Toxicity testing in the 21st century: progress in the past decade and future perspectives	Advances in the biological sciences have led to an ongoing paradigm shift in toxicity testing based on expanded application of high-throughput in vitro screening and in silico methods to assess potential health risks of environmental agents. This review examines progress on the vision for toxicity testing elaborated by the US National Research Council (NRC) during the decade that has passed since the 2007 NRC report on Toxicity Testing in the 21st Century (TT21C). Concomitant advances in exposure assessment, including computational approaches and high-throughput exposomics, are also documented. A vision for the next generation of risk science, incorporating risk assessment methodologies suitable for the analysis of new toxicological and exposure data, resulting in human exposure guidelines is described. Case study prototypes indicating how these new approaches to toxicity testing, exposure measurement, and risk assessment are beginning to be applied in practice are presented. Overall, progress on the 20-year transition plan laid out by the US NRC in 2007 has been substantial. Importantly, government agencies within the United States and internationally are beginning to incorporate the new approach methodologies envisaged in the original TT21C vision into regulatory practice. Future perspectives on the continued evolution of toxicity testing to strengthen regulatory risk assessment are provided.	D. Krewski, M. E. Andersen, M. G. Tyshenko, K. Krishnan, T. Hartung, K. Boekelheide, J. F. Wambaugh, D. Jones, M. Whelan, R. Thomas, C. Yauk, T. Barton-Maclaren & I. Cote	Archives of Toxicology			2020	Si
210	Springer Link	The Effects of Mindfulness Meditation on Attention, Executive Control and Working Memory in Healthy Adults: A Meta-analysis of Randomized Controlled Trials	Mindfulness based interventions (MBI) are becoming increasingly popular. Given their nature (i.e., training of focused attention and cognitive control), efforts have been made to study their potential benefits to different aspects of cognition, resulting in mixed results. In light of the inconsistent findings, concerns regarding the methodological quality of such studies, and recent surge in randomized controlled trials of mindfulness interventions, we conducted a meta-analysis focused on MBIs effects on attention, working-memory and executive control in healthy adults.	Ofir Yakobi, Daniel Smilek & James Danckert	Cognitive Therapy and Research			2021	Si
211	Springer Link	COVID-19 transmission during swimming-related activities: a rapid systematic review	There are uncertainties about mitigating strategies for swimming-related activities in the context of the COVID-19 pandemic. There is an opportunity to learn from the experience of previous re-openings to better plan the future one. Our objectives are to systematically review the evidence on (1) the association between engaging in swimming-related activities and COVID-19 transmission; and (2) the effects of strategies for preventing COVID-19 transmission during swimming-related activities.	Sally Yaacoub, Joanne Khabsa, Rayane El-Khoury, Amena El-Harakeh, Tamara	BMC Infectious Diseases			2021	Si

				Lotfi, Zahra Saad, Zeina Itani, Assem M. Khamis, Ibrahim El Mikati, Carlos A. Cuello-Garcia, Francisca Verdugo-Paiva, Gabriel Rada, Holger J. Schünemann, Nesrine Rizk & Elie A. Akl					
212	Springer Link	Improving reference prioritisation with PICO recognition	Machine learning can assist with multiple tasks during systematic reviews to facilitate the rapid retrieval of relevant references during screening and to identify and extract information relevant to the study characteristics, which include the PICO elements of patient/population, intervention, comparator, and outcomes. The latter requires techniques for identifying and categorising fragments of text, known as named entity recognition.	Austin J. Brockmeier, Meizhi Ju, Piotr Przybyła & Sophia Ananiadou	BMC Medical Informatics and Decision Making			2019	Si
213	Springer Link	OpenTrials: towards a collaborative open database of all available information on all clinical trials	OpenTrials is a collaborative and open database for all available structured data and documents on all clinical trials, threaded together by individual trial. With a versatile and expandable data schema, it is initially designed to host and match the following documents and data for each trial: registry entries; links, abstracts, or texts of academic journal papers; portions of regulatory documents describing individual trials; structured data on methods and results extracted by systematic reviewers or other researchers; clinical study reports; and additional documents such as blank consent forms, blank case report forms, and protocols. The intention is to create an open, freely re-usable index of all such information and to increase discoverability, facilitate research, identify inconsistent data, enable audits on the availability and completeness of this information, support advocacy for better data and drive up standards around open data in evidence-based medicine. The project has phase I funding. This will allow us to create a practical data schema and populate the database initially through web-scraping, basic record linkage techniques, crowd-sourced curation	Ben Goldacre & Jonathan Gray	Trials			2016	Si

			around selected drug areas, and import of existing sources of structured and documents. It will also allow us to create user-friendly web interfaces onto the data and conduct user engagement workshops to optimise the database and interface designs. Where other projects have set out to manually and perfectly curate a narrow range of information on a smaller number of trials, we aim to use a broader range of techniques and attempt to match a very large quantity of information on all trials. We are currently seeking feedback and additional sources of structured data.						
214	Springer Link	Barriers and facilitators to the implementation of orthodontic mini-implants in clinical practice: a protocol for a systematic review and meta-analysis	Most orthodontic treatment plans need some form of anchorage to control the reciprocal forces of tooth movement. Orthodontic mini implants (OMIs) have been hailed for having revolutionized orthodontics, because they provide anchorage without depending on the collaboration of patients, they have a favorable effectiveness compared with conventional anchorage devices, and they can be used for a wide scale of treatment objectives. However, surveys have shown that many orthodontists never or rarely use them. To understand the rationale behind this knowledge-to-action gap, we will conduct a systematic review that will identify and quantify potential barriers and facilitators to the implementation of OMIs in clinical practice for all potential stakeholders, i.e., patients, family members, clinicians, office staff, clinic owners, policy makers, etc. The prevalence of clinicians that do not use OMIs will be our secondary outcome.	Reint Meursinge Reynders, Laura Ronchi, Luisa Ladu, Nicola Di Girolamo, Jan de Lange, Nia Roberts & Sharon Mickan	Systematic Reviews			2016	Si
215	Springer Link	Barriers and facilitators to the implementation of orthodontic mini implants in clinical practice: a systematic review	Numerous surveys have shown that orthodontic mini implants (OMIs) are underused in clinical practice. To investigate this implementation issue, we conducted a systematic review to (1) identify barriers and facilitators to the implementation of OMIs for all potential stakeholders and (2) quantify these implementation constructs, i.e., record their prevalence. We also recorded the prevalence of clinicians in the eligible studies that do not use OMIs.	Reint Meursinge Reynders, Laura Ronchi, Luisa Ladu, Nicola Di Girolamo, Jan de Lange, Nia Roberts & Sharon Mickan	Systematic Reviews			2016	Si
216	Springer Link	The effect of combined cognitive intervention and physical exercise on cognitive function in older adults with mild cognitive impairment: a meta-analysis of randomized	The state of mild cognitive impairment (MCI) provides an optimal window for preventing progression to dementia. Combined cognitive intervention and physical exercise may yield additive and synergistic effects on cognition in older adults with MCI.	Qiuyan Meng, Huiru Yin, Shuo Wang, Binghan Shang, Xiangfei Meng,	Aging Clinical and Experimental Research			2021	Si

		controlled trials		Mingli Yan, Guichen Li, Jianfeng Chu & Li Chen					
217	Springer Link	A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss	This systematic review aimed to evaluate AI chatbot characteristics, functions, and core conversational capacities and investigate whether AI chatbot interventions were effective in changing physical activity, healthy eating, weight management behaviors, and other related health outcomes.	Yoo Jung Oh, Jingwen Zhang, Min-Lin Fang & Yoshimi Fukuoka	International Journal of Behavioral Nutrition and Physical Activity			2021	Si
218	Springer Link	Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools	Systematic reviews and systematic maps represent powerful tools to identify, collect, evaluate and summarise primary research pertinent to a specific research question or topic in a highly standardised and reproducible manner. Even though they are seen as the “gold standard” when synthesising primary research, systematic reviews and maps are typically resource-intensive and complex activities. Thus, managing the conduct and reporting of such reviews can become a time consuming and challenging task. This paper introduces the open access online tool CADIMA, which was developed through a collaboration between the Julius Kühn-Institut and the Collaboration for Environmental Evidence, in order to increase the efficiency of the evidence synthesis process and facilitate reporting of all activities to maximise methodological rigour. Furthermore, we analyse how CADIMA compares with other available tools by providing a comprehensive summary of existing software designed for the purposes of systematic review management. We show that CADIMA is the only available open access tool that is designed to: (1) assist throughout the systematic review/map process; (2) be suited to reviews broader than medical sciences; (3) allow for offline data extraction; and, (4) support working as a review team.	Christian Kohl, Emma J. McIntosh, Stefan Unger, Neal R. Haddaway, Steffen Kecke, Joachim Schiemann & Ralf Wilhelm	Environmental Evidence			2018	Si
219	Springer Link	Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR)	Systematic reviews (SR) are vital to health care, but have become complicated and time-consuming, due to the rapid expansion of evidence to be synthesised. Fortunately, many tasks of systematic reviews have the potential to be automated or may be assisted by automation. Recent advances in natural language processing, text mining and machine learning have produced new algorithms that can accurately mimic human endeavour in systematic review activity, faster and more cheaply. Automation tools need to be able to work together, to exchange data and results. Therefore, we initiated the International Collaboration for the Automation of Systematic Reviews (ICASR), to successfully put all the parts of automation of systematic review production together. The first meeting was held in Vienna in	Elaine Beller, Justin Clark, Guy Tsafnat, Clive Adams, Heinz Diehl, Hans Lund, Mourad	Systematic Reviews			2018	Si

			October 2015. We established a set of principles to enable tools to be developed and integrated into toolkits. This paper sets out the principles devised at that meeting, which cover the need for improvement in efficiency of SR tasks, automation across the spectrum of SR tasks, continuous improvement, adherence to high quality standards, flexibility of use and combining components, the need for a collaboration and varied skills, the desire for open source, shared code and evaluation, and a requirement for replicability through rigorous and open evaluation. Automation has a great potential to improve the speed of systematic reviews. Considerable work is already being done on many of the steps involved in a review. The 'Vienna Principles' set out in this paper aim to guide a more coordinated effort which will allow the integration of work by separate teams and build on the experience, code and evaluations done by the many teams working across the globe.	Ouzzani, Kristina Thayer, James Thomas, Tari Turner, Jun Xia, Karen Robinson & Paul Glasziou On behalf of the founding members of the ICASR group					
220	Springer Link	Comparison of minimally invasive percutaneous fixation and open reduction internal fixation for patella fractures: a meta-analysis	Open reduction internal fixation (ORIF) has long been the conventional procedure for managing displaced patella fracture. This surgical approach has certain drawbacks, which might affect clinical outcomes and patient prognosis. Minimally invasive percutaneous fixation (MIPF) was proposed to overcome these disadvantages. Few in-depth investigations have been performed to determine the superiority of MIPF over ORIF. The aim of this study was to compare the efficacies of MIPF and ORIF for patella fractures.	Chun-Hong Lo & Chih-Hwa Chen	Journal of Orthopaedic Surgery and Research			2021	Si
221	Springer Link	Big Data-Enabled Nursing	The practice of using discrete, quantitative data from multiple data sources and evidence to support clinical decisions began centuries ago when Florence Nightingale invented polar-area diagrams to show that many army soldiers' deaths could be traced to unsanitary clinical practices, and therefore, were preventable. Today, the massive volume of healthcare data that is generated from our healthcare systems requires sophisticated information technologies for storing, aggregation and analyses. Additionally the complexities of data generated from unconnected, disparate systems present challenges because new data science and big data analytics methods must be used to unlock their potential for providing answers. The nurse's unique challenge is to make sense of all the data coming from disparate sources and derive useful actionable information. This chapter and its accompanying four case studies address how the everyday use of information generated through big data analytics can inform practice for frontline nurses: for effectiveness and quality improvement; for changing care delivery that requires sharing information within a care team and across settings; and, for academic researchers for the generation of new knowledge and science.	Ellen M. HarperEmail authorDouglas McNair					2017	Si

222	Springer Link	Lifting cognition: a meta-analysis of effects of resistance exercise on cognition	The health benefits of resistance exercises are well established; however, the effects of resistance training on cognition are not as well understood. The purpose of this meta-analysis was to evaluate the evidence of resistance exercise's effects on cognition. A systematic search identified 24 studies that were included in the analyses. These articles ranged in the protocols utilized and in how they studied the effects of resistance training on cognition. Four primary analyses were carried out to assess the effects of resistance exercise on cognitive outcomes: (1) composite cognitive scores, (2) screening measures of cognitive impairment, (3) measures of executive functions, and (4) measures of working memory. Results revealed positive effects of resistance training on composite cognitive scores (SMD 0.71, 95% CI 0.30–1.12), screening measures of cognitive impairment (SMD 1.28, 95% CI 0.39–2.18), and executive functions (SMD 0.39, 95% CI 0.04–0.74), but no effect on measures of working memory (SMD 0.151, 95% CI − 0.21 to 0.51). High heterogeneity was observed in all analyses. Resistance training appears to have positive effects on cognition; however, future research will need to determine why the effects are so variable.	Jon-Frederick Landrigan, Tyler Bell, Michael Crowe, Olivio J. Clay & Daniel Mirman	Psychological Research			2020	Si
223	Springer Link	Strategies for improving the lives of US women aged 40 and above living with HIV/AIDS: an evidence map	While in its early years the HIV epidemic affected primarily the male and the young, nowadays, the population living with HIV/AIDS is approximately 24% women, and its age composition has shifted towards older ages. Many of the older women who live with HIV/AIDS also live with the medical and social conditions that accompany aging. This work aims to identify and characterize empirical studies of strategies for the comprehensive management of women over 40, including transgender women, who live with HIV/AIDS. Forty was chosen as an operational age cutoff to identify premenopausal women who are less likely to bear children, as well as peri- and postmenopausal women.	Gaelen P. Adam, Mengyang Di, Susan Cu-Uvin, Christopher Halladay, Bryant T. Smith, Suchitra Iyer & Thomas A. Trikalinos	Systematic Reviews			2018	Si
224	Medline	Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies	To systematically examine the design, reporting standards, risk of bias, and claims of studies comparing the performance of diagnostic deep learning algorithms for medical imaging with that of expert clinicians.	Myura Nagendran 1, Yang Chen 2, Christopher A Lovejoy 3, Anthony C Gordon 4 5, Matthieu Komorowski 6, Hugh Harvey 7, Eric J	BMJ	Imperial College London, University College London,	Estados unidos, Reino Unido	2020	Si

				Topol 8, John P A Ioannidis 9, Gary S Collins 10 11, Mahiben Maruthappu					
225	Medline	Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study	Colonoscopy with computer-aided detection (CADe) has been shown in non-blinded trials to improve detection of colon polyps and adenomas by providing visual alarms during the procedure. We aimed to assess the effectiveness of a CADe system that avoids potential operational bias.	Pu Wang 1, Xiaogang Liu 1, Tyler M Berzin 2, Jeremy R Glissen Brown 2, Peixi Liu 1, Chao Zhou 1, Lei Lei 1, Liangping Li 1, Zhenzhen Guo 1, Shan Lei 1, Fei Xiong 1, Han Wang 1, Yan Song 1, Yan Pan 1, Guanyu Zhou	Lancet Gastroenterol Hepatol	Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Beth Israel Deaconess Medical Center and Harvard Medical School	China, Estados Unidos	2020	no
226	Medline	Mycophenolic Acid Exposure Prediction Using Machine Learning	Therapeutic drug monitoring of mycophenolic acid (MPA) based on area under the curve (AUC) is well-established and machine learning (ML) approaches could help to estimate AUC. The aim of this work is to estimate the AUC of MPA in organ transplant patients using extreme gradient boosting (Xgboost R package) ML models. A total of 12,877 MPA AUC from 0 to 12 hours (AUC0-12 h) requests from 6,884 patients sent to our Immunosuppressant Bayesian Dose Adjustment expert system (https://abis.chu-limoges.fr) for AUC estimation and dose recommendation based on MPA concentrations measured at least at three sampling times (~ 20 minutes, 1 and 3 hours after dosing) were used to develop two ML models based on two or three concentrations. Data were split into a training set (75%) and a test set (25%) and the Xgboost models in the training set with	Jean-Baptiste Woillard 1 2, Marc Labriffe 1 2, Jean Debord 1 2, Pierre Marquet	Clin Pharmacol Ther	Université de Limoges,	Francia	2021	no

			the lowest root mean squared error (RMSE) in a 10-fold cross-validation experiment were evaluated in the test set and in 4 independent full-pharmacokinetic (PK) datasets from renal or heart transplant recipients. ML models based on two or three concentrations, differences between these concentrations, relative deviations from theoretical times of sampling, presence of a delayed absorption peak, and five covariates (dose, type of transplantation, associated immunosuppressant, age, and time between transplantation and sampling) yielded accurate AUC estimation performances in the test datasets (relative bias < 5% and relative RMSE < 20%) and better performance than MAP Bayesian estimation in the four independent full-PK datasets. The Xgboost ML models described allow accurate estimation of MPA AUC0-12 h and can be used for routine exposure estimation and dose adjustment and will soon be implemented in a dedicated web interface.						
227	Medline	Predictive performance and impact of algorithms in remote monitoring of chronic conditions: A systematic review and meta-analysis	The use of telehealth interventions, such as the remote monitoring of patient clinical data (e.g. blood pressure, blood glucose, heart rate, medication use), has been proposed as a strategy to better manage chronic conditions and to reduce the impact on patients and healthcare systems. The use of algorithms for data acquisition, analysis, transmission, communication and visualisation are now common in remote patient monitoring. However, their use and impact on chronic disease management has not been systematically investigated.	Grant Castelyn, Liliana Laranjo, Günter Schreier, Blanca Gallego	Int J Med Inform	Macquarie University, The University of Sydney, University of New South Wales	Australia, Austria	2021	Si
228	Medline	Central Reading of Ulcerative Colitis Clinical Trial Videos Using Neural Networks	Endoscopic disease activity scoring in ulcerative colitis (UC) is useful in clinical practice but done infrequently. It is required in clinical trials, where it is expensive and slow because human central readers are needed. A machine learning algorithm automating the process could elevate clinical care and facilitate clinical research. Prior work using single-institution databases and endoscopic still images has been promising.	Klaus Gottlieb, James Requa, William Karnes, Ranga Chandra Gudivada, Jie Shen, Efren Rael, Vipin Arora, Tyler Dao, Andrew Ninh, James McGill	Gastroenterology			2021	Si
229	Medline	Machine learning to help researchers evaluate biases in clinical trials: a prospective, randomized user	Assessing risks of bias in randomized controlled trials (RCTs) is an important but laborious task when conducting systematic reviews. RobotReviewer (RR), an open-source machine learning (ML) system, semi-automates bias assessments. We conducted a user study of RobotReviewer, evaluating time saved and usability of the tool.	Frank Soboczenski, Thomas A Trikalinos, Joël	BMC Med Inform Decis Mak	Brown University, University of Texas at Austin, Northeastern University, King's	Estados Unidos, Reino Unido, Holanda	2019	Si

		study		Kuiper, Randolph G Bias, Byron C Wallace, Iain J Marshall		College London			
230	Medline	Comparing machine and human reviewers to evaluate the risk of bias in randomized controlled trials	Evidence from new health technologies is growing, along with demands for evidence to inform policy decisions, creating challenges in completing health technology assessments (HTAs)/systematic reviews (SRs) in a timely manner. Software can decrease the time and burden by automating the process, but evidence validating such software is limited. We tested the accuracy of RobotReviewer, a semi-autonomous risk of bias (RoB) assessment tool, and its agreement with human reviewers.	Susan Armijo-Olivo, Rodger Craig, Sandy Campbell	Res Synth Methods	University of Alberta, University of Applied Sciences	Canada, Alemania	2020	No
231	Medline	RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials	To develop and evaluate RobotReviewer, a machine learning (ML) system that automatically assesses bias in clinical trials. From a (PDF-formatted) trial report, the system should determine risks of bias for the domains defined by the Cochrane Risk of Bias (RoB) tool, and extract supporting text for these judgments.	Iain J Marshall, Joël Kuiper, Byron C Wallace	J Am Med Inform Assoc	King's College London, University of Groningen, University of Texas at Austin	Estados Unidos, Reino Unido, Holanda	2016	Si
232	Medline	Machine learning analysis plans for randomised controlled trials: detecting treatment effect heterogeneity with strict control of type I error	Retrospective exploratory analyses of randomised controlled trials (RCTs) seeking to identify treatment effect heterogeneity (TEH) are prone to bias and false positives. Yet the desire to learn all we can from exhaustive data measurements on trial participants motivates the inclusion of such analyses within RCTs. Moreover, widespread advances in machine learning (ML) methods hold potential to utilise such data to identify subjects exhibiting heterogeneous treatment response.	James A Watson, Chris C Holmes	Trials	Mahidol University, University of Oxford	Reino Unido, Tailandia	2020	Si
233	Medline	Pharmacoepidemiology, Machine Learning, and COVID-19: An Intent-to-Treat Analysis of Hydroxychloroquine, With or Without Azithromycin, and COVID-19 Outcomes Among Hospitalized US Veterans	Hydroxychloroquine (HCQ) was proposed as an early therapy for coronavirus disease 2019 (COVID-19) after in vitro studies indicated possible benefit. Previous in vivo observational studies have presented conflicting results, though recent randomized clinical trials have reported no benefit from HCQ among patients hospitalized with COVID-19. We examined the effects of HCQ alone and in combination with azithromycin in a hospitalized population of US veterans with COVID-19, using a propensity score-adjusted survival analysis with imputation of missing data. According to electronic health record data from the US Department of Veterans Affairs health care system, 64,055 US Veterans were tested for the virus that causes COVID-19 between March 1, 2020 and April 30, 2020. Of the 7,193 veterans who tested positive, 2,809 were hospitalized, and 657 individuals were prescribed HCQ within the first 48-hours of hospitalization for the treatment of COVID-19. There was no apparent benefit associated with HCQ receipt, alone or in combination with azithromycin, and there was an increased risk of intubation when HCQ was used in combination with azithromycin (hazard ratio = 1.55; 95% confidence interval: 1.07, 2.24). In	Hanna Gerlovin, Daniel C Posner, Yuk-Lam Ho, Christopher T Rentsch, Janet P Tate, Joseph T King, Katherine E Kurgansky, Ioana Danciu, Lauren	Am J Epidemiol			2021	Si

			conclusion, we assessed the effectiveness of HCQ with or without azithromycin in treatment of patients hospitalized with COVID-19, using a national sample of the US veteran population. Using rigorous study design and analytic methods to reduce confounding and bias, we found no evidence of a survival benefit from the administration of HCQ.	Costa, Franciel A Linares, Ian D Goethert, Daniel A Jacobson, Matthew S Freiberg, Edmon Begoli, Sumitra Muralidhar, Rachel B Ramoni, Georgia Tourassi, J Michael Gaziano, Amy C Justice, David R Gagnon, Kelly Cho					
234	Medline	Physical Activity Change in an RCT: Comparison of Measurement Methods	We aimed to quantify the agreement between self-report, standard cut-point accelerometer, and machine learning accelerometer estimates of physical activity (PA), and exam- ine how agreement changes over time among older adults in an intervention setting. Methods: Data were from a randomized weight loss trial that encouraged increased PA among 333 postmenopausal breast cancer survivors. PA was estimated using accelerometry and a validated questionnaire at baseline and 6-months. Accelerometer data were processed using standard cut-points and a validated machine learning algorithm. Agreement of PA at each time-point and change was assessed using mixed effects regression models and concordance correlation. Results: At baseline, self-report and machine learning provided similar PA estimates (mean dif- ference = 11.5 min/day) unlike self-report and standard cut-points (mean difference = 36.3 min/ day). Cut-point and machine learning methods assessed PA change over time more similarly than other comparisons. Specifically, the mean difference of PA change for the cut-point versus machine learning methods was 5.1 min/day for intervention group and 2.9 in controls, whereas it was ≥ 24.7 min/day for other comparisons. Conclusions: Intervention researchers are facing the issue of self-report measures introducing bias and accelerometer cut-points being insensi- tive. Machine learning approaches may bridge this gap.	Sandahl H Nelson, Loki Natarajan, Ruth E Patterson, Sheri J Hartman, Caroline A Thompson, Suneeta V Godbole, Eileen Johnson, Catherine R Marinac, Jacqueline Kerr	Am J Health Behav			2019	Si

235	Medline	Radiomic model for differentiating parotid pleomorphic adenoma from parotid adenolymphoma based on MRI images	Distinguishing parotid pleomorphic adenoma (PPA) from parotid adenolymphoma (PA) is important for precision treatment, but there is a lack of readily available diagnostic methods. In this study, we aimed to explore the diagnostic value of radiomic signatures based on magnetic resonance imaging (MRI) for PPA and PA.	Le-le Song, Shun-Jun Chen, Wang Chen, Zhan Shi, Xiao-Dong Wang, Li-Na Song, Dian-Sen Chen	BMC Med Imaging			2021	Si
236	Medline	Physical characteristics not psychological state or trait characteristics predict motion during resting state fMRI	Head motion (HM) during fMRI acquisition can significantly affect measures of brain activity or connectivity even after correction with preprocessing methods. Moreover, any systematic relationship between HM and variables of interest can introduce systematic bias. There is a large and growing interest in identifying neural biomarkers for psychiatric disorders using resting state fMRI (rsfMRI). However, the relationship between HM and different psychiatric symptoms domains is not well understood. The aim of this investigation was to determine whether psychiatric symptoms and other characteristics of the individual predict HM during rsfMRI. A sample of n = 464 participants (174 male) from the Tulsa1000, a naturalistic longitudinal study recruiting subjects with different levels of severity in mood/anxiety/substance use disorders based on the dimensional NIMH Research Domain Criteria framework was used for this study. Based on a machine learning (ML) pipeline with nested cross-validation to avoid overfitting, the stacked model with 15 anthropometric (like body mass index, BMI) and demographic (age and sex) variables identifies BMI and weight as the most important variables and explained 10.9 percent of the HM variance (95% CI: 9.9-11.8). In comparison ML models with 105 self-report measures for state and trait psychological characteristics identified nicotine and alcohol use variables as well as impulsivity inhibitory control variables but explain only 5 percent of HM variance (95% CI: 3.5-6.4). A combined ML model using all 120 variables did not perform significantly better than the model using only 15 physical variables (combined model 95% confidence interval: 10.2-12.4). Taken together, after considering physical variables, state or trait psychological characteristics do not provide additional power to predict motion during rsfMRI.	Hamed Ekhtiari, Rayus Kuplicki, Hung-Wen Yeh, Martin P Paulus	Sci Rep			2019	Si

237	Medline	Application of Machine Learning for Tumor Growth Inhibition - Overall Survival Modeling Platform	Machine learning (ML) was used to leverage tumor growth inhibition (TGI) metrics to characterize the relationship with overall survival (OS) as a novel approach and to compare with traditional TGI-OS modeling methods. Historical dataset from a phase III non-small cell lung cancer study (OAK, atezolizumab vs. docetaxel, N = 668) was used. ML methods support the validity of TGI metrics in predicting OS. With lasso, the best model with TGI metrics outperforms the best model without TGI metrics. Boosting was the best linear ML method for this dataset with reduced estimation bias and lowest Brier score, suggesting better prediction accuracy. Random forest did not outperform linear ML methods despite hyperparameter optimization. Kernel machine was marginally the best nonlinear ML method for this dataset and uncovered nonlinear and interaction effects. Nonlinear ML may improve prediction by capturing nonlinear effects and covariate interactions, but its predictive performance and value need further evaluation with larger datasets.	Phyllis Chan, Xiaofei Zhou, Nina Wang, Qi Liu, René Bruno, Jin Y Jin	CPT Pharmacometrics Syst Pharmacol	The Ohio State University	Estados Unidos, Francia	2021	Si
238	Medline	A Convolutional Neural Network for Real Time Classification, Identification, and Labelling of Vocal Cord and Tracheal Using Laryngoscopy and Bronchoscopy Video	The use of artificial intelligence, including machine learning, is increasing in medicine. Use of machine learning is rising in the prediction of patient outcomes. Machine learning may also be able to enhance and augment anesthesia clinical procedures such as airway management. In this study, we sought to develop a machine learning algorithm that could classify vocal cords and tracheal airway anatomy real-time during video laryngoscopy or bronchoscopy as well as compare the performance of three novel convolutional networks for detecting vocal cords and tracheal rings.	Clyde Matava, Evelina Pankiv, Sam Raisbeck, Monica Caldeira, Fahad Alam	J Med Syst			2020	Disponible en Springer Link

239	Medline	Agreement in Risk of Bias Assessment Between RobotReviewer and Human Reviewers: An Evaluation Study on Randomised Controlled Trials in Nursing-Related Cochrane Reviews	RobotReviewer is a machine learning system for semi-automated assistance in risk of bias assessment. The tools's performance in randomized controlled trials (RCTs) in the field of nursing remains unknown. We aimed therefore to evaluate the agreement in risk of bias assessment between RobotReviewer and human reviewers.	Julian Hirt, Jasmin Meichlinger, Petra Schumacher, Gerhard Mueller	J Nurs Scholarsh	University for Applied Sciences FHS, Martin Luther University Halle-Wittenberg, IMC University of Applied Sciences Krems, UMIT-Private University for Health Sciences	Austria, Alemania, Suiza	2021	No
240	Medline	Diagnosis and risk stratification in hypertrophic cardiomyopathy using machine learning wall thickness measurement: a comparison with human test-retest performance	Left ventricular maximum wall thickness (MWT) is central to diagnosis and risk stratification of hypertrophic cardiomyopathy, but human measurement is prone to variability. We developed an automated machine learning algorithm for MWT measurement and compared precision (reproducibility) with that of 11 international experts, using a dataset of patients with hypertrophic cardiomyopathy.	João B Augusto, Rhodri H Davies, Anish N Bhuva, Kristopher D Knott, Andreas Seraphim, Mashael Alfarih, Clement Lau, Rebecca K Hughes, Luís R Lopes, Hunain Shiwani, Thomas A Treibel, Bernhard L Gerber, Christian Hamilton-Craig, Ntobeko A B Ntusi, Gianluca	Lancet Digit Health			2021	Disponible en Science Direct

				Pontone, Milind Y Desai, John P Greenwood, Peter P Swoboda, Gabriella Captur, João Cavalcante, Chiara Bucciarelli-Ducci, Steffen E Petersen, Erik Schelbert, Charlotte Manisty, James C Moon					
241	Medline	AIPW: An R Package for Augmented Inverse Probability-Weighted Estimation of Average Causal Effects	An increasing number of recent studies have suggested that doubly robust estimators with cross-fitting should be used when estimating causal effects with machine learning methods. However, not all existing programs that implement doubly robust estimators support machine learning methods and cross-fitting, or provide estimates on multiplicative scales. To address these needs, we developed AIPW, a software package implementing augmented inverse probability weighting (AIPW) estimation of average causal effects in R (R Foundation for Statistical Computing, Vienna, Austria). Key features of the AIPW package include cross-fitting and flexible covariate adjustment for observational studies and randomized controlled trials (RCTs). In this paper, we use a simulated RCT to illustrate implementation of the AIPW estimator. We also perform a simulation study to evaluate the performance of the AIPW package compared with other doubly robust implementations, including CausalGAM, npcausal, tmle, and tmle3. Our simulation showed that the AIPW package yields performance comparable to that of other programs. Furthermore, we also found that cross-fitting substantively decreases the bias and improves the confidence interval coverage for doubly robust estimators fitted with machine learning algorithms. Our findings suggest that the AIPW package can be a useful tool for estimating average causal effects with machine learning methods in RCTs and observational studies.	Yongqi Zhong, Edward H Kennedy, Lisa M Bodnar, Ashley I Naimi	Am J Epidemiol			2021	Si
242	Medline	Technology-assisted risk of bias assessment in	To evaluate the reliability of RobotReviewer's risk of bias judgments.	Allison Gates, Ben	J Clin Epidemiol	University of Alberta	Canada	2018	Si

		systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool		Vandermeer, Lisa Hartling					
243	Medline	Development of attenuation correction methods using deep learning in brain-perfusion single-photon emission computed tomography	Computed tomography (CT)-based attenuation correction (CTAC) in single-photon emission computed tomography (SPECT) is highly accurate, but it requires hybrid SPECT/CT instruments and additional radiation exposure. To obtain attenuation correction (AC) without the need for additional CT images, a deep learning method was used to generate pseudo-CT images has previously been reported, but it is limited because of cross-modality transformation, resulting in misalignment and modality-specific artifacts. This study aimed to develop a deep learning-based approach using non-attenuation-corrected (NAC) images and CTAC-based images for training to yield AC images in brain-perfusion SPECT. This study also investigated whether the proposed approach is superior to conventional Chang's AC (ChangAC).	Taisuke Murata, Hajime Yokota, Ryuhei Yamato, Takuro Horikoshi, Masato Tsuneda, Ryuna Kurosawa, Takuma Hashimoto, Joji Ota, Koichi Sawada, Takashi Iimori, Yoshitada Masuda, Yasukuni Mori, Hiroki Suyari, Takashi Uno	Med Phys	Chiba University	Japan	2021	No
244	Medline	Added Value of Deep Learning-based Detection System for Multiple Major Findings on Chest Radiographs: A Randomized Crossover Study	Background Previous studies assessing the effects of computer-aided detection on observer performance in the reading of chest radiographs used a sequential reading design that may have biased the results because of reading order or recall bias. Purpose To compare observer performance in detecting and localizing major abnormal findings including nodules, consolidation, interstitial opacity, pleural effusion, and pneumothorax on chest radiographs without versus with deep learning-based detection (DLD) system assistance in a randomized crossover design. Materials and Methods This study included retrospectively collected normal and abnormal chest radiographs between January 2016 and December 2017 (https://cris.nih.go.kr/; registration no. KCT0004147). The radiographs were randomized into two groups, and six observers, including thoracic radiologists, interpreted each radiograph without and with use of a commercially available DLD system by using a crossover design with a washout period. Jackknife alternative free-	Jinkyeong Sung, Sohee Park, Sang Min Lee, Woong Bae 1, Beomhee Park, Eunkyung Jung, Joon Beom Seo, Kyu-Hwan	Radiology	University of Ulsan College of Medicine	Corea del Sur	2021	No

			response receiver operating characteristic (JAFROC) figure of merit (FOM), area under the receiver operating characteristic curve (AUC), sensitivity, specificity, false-positive findings per image, and reading times of observers with and without the DLD system were compared by using McNemar and paired t tests. Results A total of 114 normal (mean patient age ± standard deviation, 51 years ± 11; 58 men) and 114 abnormal (mean patient age, 60 years ± 15; 75 men) chest radiographs were evaluated. The radiographs were randomized to two groups: group A (n = 114) and group B (n = 114). Use of the DLD system improved the observers' JAFROC FOM (from 0.90 to 0.95, P = .002), AUC (from 0.93 to 0.98, P = .002), per-lesion sensitivity (from 83% [822 of 990 lesions] to 89.1% [882 of 990 lesions], P = .009), per-image sensitivity (from 80% [548 of 684 radiographs] to 89% [608 of 684 radiographs], P = .009), and specificity (from 89.3% [611 of 684 radiographs] to 96.6% [661 of 684 radiographs], P = .01) and reduced the reading time (from 10-65 seconds to 6-27 seconds, P < .001). The DLD system alone outperformed the pooled observers (JAFROC FOM: 0.96 vs 0.90, respectively, P = .007; AUC: 0.98 vs 0.93, P = .003). Conclusion Observers including thoracic radiologists showed improved performance in the detection and localization of major abnormal findings on chest radiographs and reduced reading time with use of a deep learning-based detection system. © RSNA, 2021 Online supplemental material is available for this article.	Jung					
245	Medline	A method to combine target volume data from 3D and 4D planned thoracic radiotherapy patient cohorts for machine learning applications	The gross tumour volume (GTV) is predictive of clinical outcome and consequently features in many machine-learned models. 4D-planning, however, has prompted substitution of the GTV with the internal gross target volume (iGTV). We present and validate a method to synthesise GTV data from the iGTV, allowing the combination of 3D and 4D planned patient cohorts for modelling.	Corinne Johnson, Gareth Price, Jonathan Khalifa, Corinne Faivre-Finn, Andre Dekker, Christopher Moore, Marcel van Herk	Radiother Oncol	The University of Manchester, Institut Universitaire du Cancer de Toulouse - Oncopole, Maastricht University Medical Centre	Holanda, Reino Unido, Francia	2018	Disponible
246	Medline	Automatic Measurement of Kidney and Liver Volumes from MR Images of Patients Affected by Autosomal Dominant Polycystic Kidney Disease	The formation and growth of cysts in kidneys, and often liver, in autosomal dominant polycystic kidney disease (ADPKD) cause progressive increases in total kidney volume (TKV) and liver volume (TLV). Laborious and time-consuming manual tracing of kidneys and liver is the current gold standard. We developed a fully automated segmentation method for TKV and TLV measurement that uses a deep learning network optimized to perform semantic segmentation of kidneys and liver.	Maatje D A van Gastel, Marie E Edwards, Vicente E Torres, Bradley J Erickson,	J Am Soc Nephrol	University of Groningen	Holanda	2019	Si

				Ron T Gansevoort, Timothy L Kline					
247	Medline	Correcting Classifiers for Sample Selection Bias in Two-Phase Case-Control Studies	Epidemiological studies often utilize stratified data in which rare outcomes or exposures are artificially enriched. This design can increase precision in association tests but distorts predictions when applying classifiers on nonstratified data. Several methods correct for this so-called sample selection bias, but their performance remains unclear especially for machine learning classifiers. With an emphasis on two-phase case-control studies, we aim to assess which corrections to perform in which setting and to obtain methods suitable for machine learning techniques, especially the random forest. We propose two new resampling-based methods to resemble the original data and covariance structure: stochastic inverse-probability oversampling and parametric inverse-probability bagging. We compare all techniques for the random forest and other classifiers, both theoretically and on simulated and real data. Empirical results show that the random forest profits from only the parametric inverse-probability bagging proposed by us. For other classifiers, correction is mostly advantageous, and methods perform uniformly. We discuss consequences of inappropriate distribution assumptions and reason for different behaviors between the random forest and other classifiers. In conclusion, we provide guidance for choosing correction methods when training classifiers on biased samples. For random forests, our method outperforms state-of-the-art procedures if distribution assumptions are roughly fulfilled. We provide our implementation in the R package sambia.	Norbert Krautenbacher, Fabian J Theis, Christiane Fuchs	Comput Math Methods Med	Technische Universität München	Alemania	2017	Si
248	Medline	Aggregator: a machine learning approach to identifying MEDLINE articles that derive from the same underlying clinical trial	It is important to identify separate publications that report outcomes from the same underlying clinical trial, in order to avoid over-counting these as independent pieces of evidence.	Weixiang Shao, Clive E Adams, Aaron M Cohen, John M Davis, Marian S McDonagh, Sujata Thakurta, Philip S Yu, Neil R Smalheiser	Methods	University of Illinois at Chicago, University of Nottingham, Oregon Health & Science University, University of Illinois at Chicago,	Reino Unido, Estados Unidos	2015	Si
249	Medline	New Automatic Tools to Identify Responders to Cardiac	New echocardiographic parameters (apical rocking [AR], septal flash [SF]) are intended to detect patterns specific to responders to cardiac resynchronization therapy (CRT). The patterns are visually recognized and qualitatively described, requiring experience and	Razvan O Mada, Peter Lysyansky	J Am Soc Echocardiogr	Catholic University Leuven	Belgica, Rumania, Alemania	2016	Si

		Resynchronization Therapy	training. Speckle-tracking echocardiography can reflect SF and AR by using newly developed, dedicated parameters, such as start systolic index (SSI) and peak longitudinal displacement (PLD). The aim of this study was to investigate whether SSI and PLD can identify potential CRT responders.	, Jürgen Duchenne, Ruxandra Beyer, Cristina Mada, Lucian Muresan, Horia Rosianu, Adela Serban, Stefan Winter, Wolfgang Fehske, Ivan Stankovic, Jens-Uwe Voigt					
250	Medline	A semi-automated machine-learning based workflow for ellipsoid zone analysis in eyes with macular edema: SCORE2 pilot study	To develop a semi-automated, machine-learning based workflow to evaluate the ellipsoid zone (EZ) assessed by spectral domain optical coherence tomography (SD-OCT) in eyes with macular edema secondary to central retinal or hemi-retinal vein occlusion in SCORE2 treated with anti-vascular endothelial growth factor agents.	Tyler Etheridge, Ellen T A Dobson, Marcel Wiedenmann, Chandana Papudesu, Ingrid U Scott, Michael S Ip, Kevin W Eliceiri, Barbara A Blodi, Amitha Domalpally	PLoS One	University of Wisconsin School of Medicine and Public Health, University of Wisconsin-Madison,	Estados Unidos	2020	Si
251	Medline	New technique for prostate volume assessment	The prostate-specific antigen density (PSAD) helps distinguish between benign prostatic hyperplasia (BPH) and prostate cancer. Accurate prostate volume (PV) assessment is necessary for PSAD calculation and both BPH diagnosis and treatment response monitoring; therefore, accurate PV measurement is increasingly becoming an essential step in the urology.	Mohamad Habes, Jeanette Bahr, Thilo Schiller, Jens-Peter Kühn, Laura Hoppe,	World J Urol	University of Greifswald	Alemania	2014	Si

				Martin Burchardt, Wolfgang Hoffmann					
252	Medline	Testing the predictive value of peripheral gene expression for nonremission following citalopram treatment for major depression	Major depressive disorder (MDD) in general, and anxious-depression in particular, are characterized by poor rates of remission with first-line treatments, contributing to the chronic illness burden suffered by many patients. Prospective research is needed to identify the biomarkers predicting nonremission prior to treatment initiation. We collected blood samples from a discovery cohort of 34 adult MDD patients with co-occurring anxiety and 33 matched, nondepressed controls at baseline and after 12 weeks (of citalopram plus psychotherapy treatment for the depressed cohort). Samples were processed on gene arrays and group differences in gene expression were investigated. Exploratory analyses suggest that at pretreatment baseline, nonremitting patients differ from controls with gene function and transcription factor analyses potentially related to elevated inflammation and immune activation. In a second phase, we applied an unbiased machine learning prediction model and corrected for model-selection bias. Results show that baseline gene expression predicted nonremission with 79.4% corrected accuracy with a 13-gene model. The same gene-only model predicted nonremission after 8 weeks of citalopram treatment with 76% corrected accuracy in an independent validation cohort of 63 MDD patients treated with citalopram at another institution. Together, these results demonstrate the potential, but also the limitations, of baseline peripheral blood-based gene expression to predict nonremission after citalopram treatment. These results not only support their use in future prediction tools but also suggest that increased accuracy may be obtained with the inclusion of additional predictors (eg, genetics and clinical scales).	Jean-Philippe Guilloux, Sabrina Bassi, Ying Ding, Chris Walsh, Gustavo Turecki, George Tseng, Jill M Cyranowski, Etienne Sibille	Neuropsychopharmacology	University of Pittsburgh	Estados Unidos	2015	Si
253	Medline	Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: a pilot study	To train a generic deep learning software (DLS) to classify breast cancer on ultrasound images and to compare its performance to human readers with variable breast imaging experience.	Anton S Becker, Michael Mueller, Elina Stoffel, Magda Marcon, Soleen Ghafoor, Andreas Boss	Br J Radiol	University Hospital of Zurich	Suiza	2018	Si
254	Medline	Radiomic features for prostate cancer detection on MRI differ between the transition and	To evaluate in a multi-institutional study whether radiomic features useful for prostate cancer (PCa) detection from 3 Tesla (T) multi-parametric MRI (mpMRI) in the transition zone (TZ) differ from those in the peripheral zone (PZ).	Shoshana B Ginsburg, Ahmad Algohary,	J Magn Reson Imaging	Case Western Reserve University, University of Turku	Estados Unidos	2017	Si

		peripheral zones: Preliminary findings from a multi-institutional study		Shivani Pahwa, Vikas Gulani, Lee Ponsky, Hannu J Aronen, Peter J Boström, Maret Böhm, Anne-Maree Haynes, Phillip Brenner, Warick Delprado, James Thompson, Marley Pulbrock, Pekka Taimen, Robert Villani, Phillip Stricker, Ardeshir R Rastinehad, Ivan Jambor, Anant Madabhushi					
255	Lilacs	Agreement in Risk of Bias Assessment Between RobotReviewer and Human Reviewers: An Evaluation Study on Randomised Controlled Trials in Nursing-Related Cochrane Reviews.	RobotReviewer is a machine learning system for semi-automated assistance in risk of bias assessment. The tools's performance in randomized controlled trials (RCTs) in the field of nursing remains unknown. We aimed therefore to evaluate the agreement in risk of bias assessment between RobotReviewer and human reviewers.	Hirt, Julian; Meichlinger, Jasmin; Schumacher, Petra; Mueller, Gerhard.	J Nurs Scholarsh	University for Applied Sciences FHS	Suiza	2021	Si
256	Lilacs	Technology-assisted risk of bias	To evaluate the reliability of RobotReviewer's risk of bias judgments. STUDY DESIGN AND	Gates, Allison;	J Clin Epidemiol	University of Alberta	Canada	2018	Si

		assessment in systematic reviews: a prospective cross-sectional evaluation of the RobotReviewer machine learning tool.		Vandermeer, Ben; Hartling, Lisa.					
257	Lilacs	Automation bias and verification complexity: a systematic review.	While potentially reducing decision errors, decision support systems can introduce new types of errors. Automation bias (AB) happens when users become overreliant on decision support, which reduces vigilance in information seeking and processing. Most research originates from the human factors literature, where the prevailing view is that AB occurs only in multitasking environments.	Lyell, David; Coiera, Enrico.	J Am Med Inform Assoc	Macquarie University	Australia	2017	Si