

Guía para la elaboración de Revisiones Sistemáticas y Metaanálisis de pruebas de diagnóstico

Un manual para el uso docente de pregrado



Guía para la elaboración de Revisiones Sistemáticas y Metaanálisis de pruebas de diagnóstico

Un manual para el uso docente de pregrado.

TM Marco Jiménez H.
Mg Epidemiología
PHD(c) Cs. Edu.
mjimenezhe@gmail.com

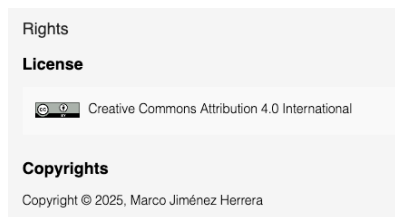
Actualización 2025

Autor:

Marco Jiménez Herrera
Diplomado en Educación en Salud
Magíster en Epidemiología
Doctorando en Ciencias de la Educación (PhD(c))
<https://orcid.org/0000-0003-2544-0275>

Santiago de Chile, 2025

© 2025 Marco Jiménez Herrera. Todos los derechos reservados.
Primera edición, Santiago de Chile, 2025.



Este manual está destinado a fines educativos y puede ser distribuido libremente para uso docente y formativo en programas de pregrado, siempre que se cite la fuente original.

Prohibida su reproducción total o parcial con fines comerciales sin autorización previa del autor.

Citación

JIMENEZ, M. (2025). Guía para la elaboración de Revisiones Sistemáticas y Metaanálisis de Pruebas de Diagnóstico. Un manual para uso docente de pregrado. Zenodo.
<https://doi.org/10.5281/zenodo.15645235>

Este manual es una guía práctica y didáctica para estudiantes de pregrado y docentes del área de la salud que se inician en el mundo de las Revisiones Sistemáticas y el Metaanálisis, particularmente en el contexto de pruebas diagnósticas. A través de un enfoque paso a paso basado en las directrices PRISMA 2020, STARD y QUADAS-2, se abordan los aspectos más relevantes para la formulación de una revisión de calidad y la interpretación crítica de la evidencia.

El autor, Marco Jiménez Herrera, es Tecnólogo Médico con más de 20 años de experiencia en el área de radiología, Magíster en Epidemiología y candidato a Doctor en Ciencias de la Educación. Actualmente se desempeña como docente universitario y coordinador de metodología de la investigación, con foco en el desarrollo de competencias en investigación aplicada en Tecnología Médica y Ciencias de la Salud.

PRESENTACIÓN DEL MANUAL

Durante la pandemia surgieron múltiples necesidades en torno a la enseñanza aprendizaje, tanto en escuelas como universidades. Un desafío que se tuvo que enfrentar en el ámbito universitario fue la continuidad de los proyectos de investigación (tesis), que se vio interrumpida por la situación sanitaria de ese momento, debiendo reformular cada proyecto en curso y en los proyectos nuevos pensar en otras alternativas que no afecten los objetivos de aprendizaje del estudiante. En este contexto surge un desafío importante, era claro que muchos proyectos debían de reconducirse hacia las revisiones de la literatura, principalmente narrativas sin rigurosidad metodológica. De aquí nace la idea de no solo presentar una revisión narrativa, si no, una revisión con características sistemáticas, que tengan una “cierta” rigurosidad en la búsqueda de información, en cuanto a criterios de elegibilidad y la forma del cómo y dónde se hace la búsqueda. Surgen comentarios como: - las revisiones están en el área de la toma de decisiones – muchas carreras de la salud no se encuentran en esa área- ¿somos capaces de guiar con éxito una revisión sistemática? Por otro lado, surgen respuestas: los profesionales del área de la salud si bien no todos están involucrados en la toma de decisiones en salud, si están involucrados con las herramientas en que se diagnostican y tratan a los pacientes, entonces, ¿no tenemos nada que decir? ¿será que los profesionales pueden mirar con ojo más crítico las pruebas de diagnóstico y así realizar un análisis crítico de la evidencia?, Pensamos que sí, los profesionales no médicos, como Tecnólogos Médicos, Enfermería, Kinesiología y muchos otros trabajan diariamente con todas las herramientas de diagnóstico, y quien más que ellos saben cuál rinde mejor que otras y que se debe de tomar en cuenta para realizar una Revisión Sistemática donde el test índice es crucial para obtener el rendimiento de una prueba.

La cuestión es si el estudiante logra identificar y realizar una Revisión Sistemática de calidad, es un desafío que no solo se lo lleva el estudiante si no el tutor quien guía el proyecto de investigación. Es aquí donde se va recogiendo la experiencia de más de 6 años realizando Revisiones Sistemáticas en pregrado, con el propósito de ir mejorando cada vez más en la calidad y en el objetivo que se persigue con el estudiante “que sea capaz de analizar

críticamente la literatura científica” y sobre eso la metodología de las Revisiones Sistemáticas.

Entonces esta guía surge de la necesidad de contar con un material claro, práctico y metodológicamente riguroso que oriente a estudiantes de pregrado y docentes del área de la salud en el desafío de comprender y elaborar Revisiones Sistemáticas con o sin Metaanálisis, especialmente centradas en pruebas diagnósticas. En mi experiencia como tecnólogo médico y formador en metodología de la investigación, he constatado las dificultades que enfrentan quienes se inician en este tipo de estudios: desde la formulación adecuada de la pregunta clínica hasta la interpretación de los resultados estadísticos.

Este manual está dirigido a:

- Estudiantes de carreras de la salud que requieren un apoyo estructurado y didáctico.
- Docentes que buscan materiales de respaldo para sus asignaturas o talleres de investigación.
- Profesionales que desean actualizar o fortalecer su comprensión sobre las revisiones sistemáticas diagnósticas.

Puede utilizarse de forma autónoma, como parte de un curso formal, o como referencia para talleres metodológicos. Cada sección está diseñada para avanzar progresivamente desde los conceptos básicos hasta las herramientas de síntesis estadística.

El lector desarrollará competencias esenciales tales como:

- Formular preguntas de investigación basadas en la estrategia PICoR.
- Aplicar criterios PRISMA 2020 y herramientas como GRADE y QUADAS-2.
- Reconocer los elementos críticos en la calidad de la evidencia diagnóstica.
- Comprender y representar gráficamente resultados de sensibilidad, especificidad y curvas ROC.

Espero que esta guía sea un instrumento valioso para la formación de competencias investigativas con impacto real en la práctica clínica y educativa.

“Cada vez que le presentamos desafíos a los estudiantes con respecto a sus temas de investigación, siempre han sabido salir adelante”

Marco Jiménez H.

Santiago, 2025

NOTA LEGAL SOBRE REUTILIZACIÓN DE CONTENIDOS

Algunas figuras y secciones metodológicas han sido adaptadas libremente con fines educativos a partir de recursos licenciados bajo Creative Commons, como PRISMA 2020 (Page et al., 2021, BMJ;372:n71) y herramientas de la Colaboración Cochrane (QUADAS-2). Todas las adaptaciones realizadas en este manual son de carácter ilustrativo, y citan adecuadamente las fuentes originales según licencia CC BY 4.0. Donde corresponda, se han creado versiones propias o recreaciones visuales con propósito pedagógico. Para más información, visite <https://www.prisma-statement.org/> y <https://www.equator-network.org/>.

TABLA DE CONTENIDOS

INTRODUCCIÓN.....	8
I. INICIANDO LA IDEA DE UNA REVISIÓN SISTEMÁTICA (RS).....	10
II. EXPLICACIÓN DEL CHECK LIST DE PRISMA 2020.....	15
PUNTO 1. TÍTULO DE LA REVISIÓN. (PICoR):	17
PUNTO 2. ABSTRACTO O RESUMEN	19
PUNTO 3. INTRODUCCIÓN.....	20
PUNTO 3. DESCRIBA LA JUSTIFICACIÓN DE LA REVISIÓN EN EL CONTEXTO DEL CONOCIMIENTO EXISTENTE.	20
PUNTO 4. OBJETIVO-PREGUNTA DE INVESTIGACIÓN.....	20
MÉTODOS.....	24
PUNTO 5. ESPECIFIQUE LOS CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN PARA LA REVISIÓN Y CÓMO SE AGRUPARON LOS ESTUDIOS PARA LAS SÍNTESIS	24
PUNTO 6. FUENTES DE INFORMACIÓN.....	26
PUNTO 7. ESTRATEGIA DE BÚSQUEDA	26
PUNTO 8. PROCESO DE SELECCIÓN.....	27
Figura 1.....	28
Diagrama Para RS de RS que incluyen búsqueda de base de datos, registros y otras fuentes o recursos.....	28
Figura 2.....	29
Diagrama para RS de RS que incluyen búsqueda de base de datos y registros.....	29
Figura 3.....	30
Diagrama para RS nuevas que incluyen búsqueda de base de datos, registros y otras fuentes o recursos.....	30
Figura 4.....	31
Diagrama para RS nuevas que incluyen búsqueda de base de datos y registros.....	31
PUNTO 9. PROCESO DE RECOPIACIÓN DE DATOS.....	32
PUNTO 10A Y 10B. ELEMENTOS DE LOS DATOS.....	32
PUNTO 11. EVALUACIÓN DEL RIESGO DEL SESGO.....	34
STARD (Standards for Reporting Diagnostic Accuracy Studies).....	35
Evaluación de sesgos y aplicabilidad según QUADAS2.....	36
Pauta QUADAS 2.....	42
PUNTO 12- MEDIDAS DE EFECTO.....	46
ODD RATIO (OR):	47
Proporciones.....	47
Precisión diagnóstica	47
Diferencia de medias.....	48
PUNTO 13 A,B,C,D,F. MÉTODOS DE SÍNTESIS	48
PUNTO 14. EVALUACIÓN SESGO DE NOTIFICACIÓN.....	49
PUNTO 15. EVALUACIÓN DE LA CERTEZA.....	50
Nivel de certeza y significado.....	51
Pirámide de la evidencia.....	51

<i>Paso a paso de la evaluación de la calidad de la evidencia según GRADE</i>	52
RESULTADOS	58
PUNTO 16A: SELECCIÓN DE ESTUDIOS:	58
PUNTO 16B: ESTUDIOS EXCLUIDOS CON JUSTIFICACIÓN	59
PUNTO 17: CARACTERÍSTICAS DE LOS ESTUDIOS INCLUIDOS	59
PUNTO 18: EVALUACIÓN DEL RIESGO DE SESGO	59
PUNTO 19: RESULTADOS DE ESTUDIOS INDIVIDUALES	60
PUNTO 20A: RESULTADOS DE SÍNTESIS (RESUMEN GENERAL)	60
PUNTO 20B: RESULTADOS ESTADÍSTICOS DEL METAANÁLISIS	60
PUNTO 20C: INVESTIGACIONES SOBRE HETEROGENEIDAD	61
PUNTO 20D: ANÁLISIS DE SENSIBILIDAD	61
PUNTO 21: SESGO DE PUBLICACIÓN (SESGOS DE INFORMACIÓN)	61
PUNTO 22: CERTEZA DE LA EVIDENCIA (GRADE)	61
EJEMPLO DE REPORTE DE RESULTADOS EN REVISIONES SISTEMÁTICAS CON METANÁLISIS ...	62
EJEMPLO DE REPORTE DE RESULTADOS EN REVISIONES SISTEMÁTICAS SIN METANÁLISIS	66
DISCUSIÓN DE UNA RS:	68
PUNTO 23A: INTERPRETACIÓN GENERAL DE LOS RESULTADOS	68
PUNTO 23B: LIMITACIONES DE LA EVIDENCIA	68
PUNTO 23C: LIMITACIONES DEL PROCESO DE REVISIÓN	68
PUNTO 23D: IMPLICANCIAS PARA LA PRÁCTICA Y FUTURAS INVESTIGACIONES	69
OTRA INFORMACIÓN RELEVANTE SEGÚN PRISMA 2020	69
PUNTO 24A: REGISTRO DE LA REVISIÓN	69
PUNTO 24B: ACCESO AL PROTOCOLO	69
PUNTO 24C: MODIFICACIONES DEL PROTOCOLO	69
PUNTO 25: APOYO FINANCIERO	69
PUNTO 26: CONFLICTO DE INTERESES	70
PUNTO 27: DISPONIBILIDAD DE DATOS Y MATERIALES	70
CONSIDERACIONES FINALES DEL AUTOR	71
ESQUEMA DE CONTENIDOS PARA REALIZAR UN RESUMEN DE UNA RS	71
RESUMEN A TRAVÉS DE PREGUNTAS Y RESPUESTAS:	72
RESUMEN FINAL PARA LA REALIZACIÓN DE RS SEGÚN PRISMA 2020	74
CONCLUSIÓN DEL MANUAL	76
REFERENCIAS	77

INTRODUCCIÓN

Las pruebas de diagnóstico permiten determinar la presencia, ausencia o severidad de una enfermedad o condición específica. En el contexto de las disciplinas propias de las ciencias de la salud y particularmente en áreas como Tecnología Médica, Imagenología, Laboratorio Clínico, Oftalmología, entre otras. Este tipo de estudios resulta especialmente relevante, ya que la mayoría de sus procedimientos constituyen pruebas diagnósticas (como la tomografía computada, radiografía, ultrasonido, resonancia magnética, análisis de laboratorio, entre otros).

Sin embargo, no todos los artículos científicos que evalúan pruebas diagnósticas presentan la calidad metodológica suficiente como para sustentar decisiones clínicas basadas en evidencia. Por esta razón, existen guías específicas orientadas a mejorar el diseño, reporte y evaluación crítica de este tipo de estudios, entre las que se destacan: STARD (para estudios primarios de test diagnóstico) (Bossuyt et al., 2015), PRISMA (para revisiones sistemáticas) (Page et al., 2021a), y QUADAS-2 (para la evaluación del riesgo de sesgo en estudios de diagnóstico) (Whiting et al., 2011). Asimismo, para estudios tipo reporte de caso, existen directrices como las pautas CARE y GRADE para evaluar la certeza de la evidencia (Balshem et al., 2011).

En el caso particular de las Revisiones Sistemáticas de pruebas diagnósticas, se recomienda utilizar el checklist de STARD para evaluar cada estudio primario incluido, mientras que la calidad metodológica global de la revisión se guía por las recomendaciones de PRISMA 2020, y la evaluación del riesgo de sesgo se apoya en la herramienta QUADAS-2, todas ellas respaldadas por la Colaboración Cochrane.

Es importante destacar que los tecnólogos médicos y otros profesionales de la salud cumplen un rol fundamental tanto en el análisis crítico de artículos científicos como en la elaboración

de Revisiones Sistemáticas de pruebas diagnósticas, debido a su experiencia especializada en el uso, interpretación y aplicación clínica de los equipos involucrados. Esta experticia resulta clave, ya que el rendimiento diagnóstico de una prueba puede variar significativamente según el tipo de equipo, protocolo técnico y condiciones operativas utilizadas. No considerar estos aspectos podría introducir sesgos importantes en una revisión sistemática, comprometiendo la validez de sus conclusiones. Por tanto, una adecuada interpretación metodológica requiere del conocimiento detallado sobre las ventajas y limitaciones de los equipos utilizados como test índice, a fin de salvaguardar la calidad de la evidencia y el impacto que esta puede tener en el diagnóstico del paciente.

El propósito de este manual es presentar de manera didáctica y aplicada las principales directrices metodológicas, en especial PRISMA 2020, para la elaboración paso a paso de una Revisión Sistemática enfocada en estudios diagnósticos. Asimismo, se destacan los elementos clave a considerar en las etapas iniciales de este tipo de investigaciones. Cabe señalar que el desarrollo de un Metaanálisis dependerá estrictamente de la calidad metodológica y homogeneidad de los estudios incluidos; en caso de estudios de baja calidad, no se recomienda su realización.

Este manual ha sido diseñado especialmente para estudiantes de pregrado y docentes del área de la salud, brindando una herramienta accesible y rigurosa para iniciarse en la lectura crítica, desarrollo y comprensión de Revisiones Sistemáticas, así como en la interpretación adecuada de Metaanálisis como instrumentos esenciales para fortalecer la toma de decisiones clínicas basadas en evidencia.

I. INICIANDO LA IDEA DE UNA REVISIÓN SISTEMÁTICA (RS)

Las RS están en el contexto de la toma de decisiones. En el escenario de tomar la decisión sobre la mejor prueba de diagnóstico para un paciente o la mejor terapia, surgen preguntas como ¿Cuál es la precisión diagnóstica de la Resonancia Magnética para diagnosticar metástasis ósea?, ¿Cuál es la precisión diagnóstica de la Tomografía Computada para detectar nódulos pulmonares?, ¿Qué elijo? Y pueden surgir muchas más preguntas, ¿entonces le preguntamos al colega con más experiencia?, hacemos una búsqueda rápida en Pubmed? Entonces podemos encontrar variabilidad en el reporte de los resultados de sensibilidad y especificidad de cada artículo científico, ¿Cuál artículo elegimos para tomar la decisión de realizar la prueba diagnóstica?

En base a esto surge la idea de realizar las revisiones sistemáticas, encontrar la mejor evidencia para resumirla y entregar resultados que permitan tomar la mejor decisión. Las RS son una síntesis de la mejor evidencia disponible para responder un objetivo o una pregunta de investigación. Se lleva a cabo mediante una búsqueda sistemática de la evidencia (de ahí su nombre) en los principales motores de búsqueda que hay en la web, como por ejemplo Pubmed, Elsevier, Cochrane, Scielo, etc. Para hacer esta búsqueda debemos de diseñar una estrategia de búsqueda con palabras claves y términos Mesh. Los términos Mesh son palabras que fueron indexadas en Pubmed y que permiten identificar ciertas palabras o conceptos, existen bastantes tutoriales de como buscar evidencia en la web. Entonces con estos conceptos podemos realizar la búsqueda sistemática de la literatura y con cierta seguridad de que no estamos excluyendo algún artículo relevante para nuestra RS. Esta búsqueda se lleva a cabo

por dos investigadores, y su búsqueda debe de ser independiente, así minimizamos el sesgo de publicación.

Si nos ponemos en la vereda de la elaboración de una RS debemos de tomar ciertas “precauciones”. ¿Es factible realizar una RS de mi tema de investigación?

Dada nuestra experiencia como docentes, los investigadores se dan cuenta muy tarde de puntos que podrían influenciar negativamente en el desarrollo de su revisión, y a veces definitivamente no poder realizarla. A continuación, se indican los puntos más relevantes a tener en cuenta en el inicio del proyecto de una RS. Más adelante lo abordaremos con mayor detalle.

- **¿Ya estará publicada nuestra idea de RS?:** Realizar una búsqueda “rápida” en los principales buscadores (Pubmed, Elsevier), con el objetivo de saber si la revisión ya está hecha. Lo otro que podemos darnos cuenta cuando revisamos la literatura es de ciertos “tips” que nos ayuden a diseñar nuestra RS.
- **¿Quiénes son los pacientes objetivo de nuestro estudio?:** Definir claramente el paciente objetivo. Hay que tener muy claro la condición del paciente, el tipo de enfermedad a estudiar. Puede ser que tengamos en mente un tipo de paciente y este sea abordado en muy pocos artículos y no sea posible realizar la RS, o una condición específica muy rara y sean muy escasas las publicaciones.
- **¿Nos sirven todo tipo de artículos científicos?:** El tipo de estudio que incluiremos en las RS es un punto muy importante y tiene que ver con el nivel de evidencia. Cuando hablamos de efectividad de una prueba o medicamento se necesitan Ensayos Clínicos Aleatorizados (ECA), estos están en la punta de la pirámide de evidencia y con un análisis crítico es el diseño correcto para incluir en nuestra revisión sistemática. Y es así cuando leemos RS, los autores declaran que es una RS de ensayos clínicos aleatorizados. En muchos casos de RS de pruebas de diagnóstico se incluyen todo tipo de estudios (ECA, caso-control, descriptivos, etc.) en estos casos lo más probable es que los resultados sean muy distintos el uno al otro y por

ello se produce una heterogeneidad muy elevada. Hablaremos de heterogeneidad de los artículos y como afecta en las RS más adelante.

- **¿Con que prueba se diagnóstica correctamente la enfermedad o condición que estamos estudiando?:** Saber cuál es la prueba de referencia con el cual se diagnóstica “con mayor precisión” es fundamental. Esta prueba es el llamado ***GOLD STANDARD O TEST DE REFERENCIA***. Es de vital importancia tomarlo en cuenta cuando se está realizando una RS que tiene como objetivo determinar la precisión diagnóstica. La prueba que estamos estudiando como por ejemplo la mamografía para determinar la precisión diagnóstica en cuanto a detectar cáncer debemos de tener en cuenta que la biopsia es el estándar de referencia, entonces debemos de procurar que todos los artículos que pretendemos incluir en nuestra RS hayan usado la biopsia como estándar de referencia.

- **¿Se necesita un metaanálisis? ¿qué datos necesito?:** El metaanálisis es un resumen estadístico de los datos que se extrajeron de los artículos seleccionados para responder los objetivos del estudio. En el caso de determinar la precisión Diagnóstica de alguna prueba (hablamos de sensibilidad y especificidad y otras pruebas de precisión) lo que necesitamos extraer de los artículos científicos son sus valores de Verdaderos Positivos (VP), Verdaderos Negativos (VN), Falsos Positivos (FP) y Falsos Negativos (FN). En determinados artículos nos encontraremos con esta información explícita, en otras ocasiones esta información debemos de deducirla de ciertos datos, por lo tanto, es fundamental si vamos a realizar un metaanálisis de los datos. Cabe señalar que el metaanálisis va a depender fundamentalmente de la heterogeneidad de los estudios y no lo sabremos hasta el final del estudio. Esta heterogeneidad puede estar dada por la variabilidad de las pruebas de referencias entre los artículos, puede ser que un artículo haya utilizado un estándar de referencia con un valor

“umbral” y otros artículos otro valor. Otro escenario de heterogeneidad es que la población de estudio difiera entre sí, las condiciones etc. Esto también ha de tenerse en cuenta en los comienzos de la realización de la RS.

Volviendo con los datos que necesitamos para el metaanálisis debemos de saber los conceptos de sensibilidad y especificidad (Cerde L & Cifuentes A, 2010)

Tabla 1.

Tabla de descripción sobre los conceptos de pruebas de diagnóstico.

Resultado de prueba de referencia (Gold Estándar)		
Test o index test (prueba a estudiar)	Presente	Ausente
Presente	VP (paciente tiene la condición confirmada por test de referencia y es positivo al index test)	FP (paciente no tiene la condición confirmada por test de referencia y el positivo al index test)
Ausente	FN (paciente tiene la condición confirmada por test de referencia y es negativo al index test) Total, de pacientes con la condición presente	VN (paciente no tiene la condición confirmada por test de referencia y es negativo al index test) Total, de pacientes con la condición ausente

De esta tabla se puede determinar la sensibilidad, especificidad, valores predictivos positivos (VPP) y negativos (VPN).

$$\text{Sensibilidad} = \frac{VP}{VP + FN} = \text{proporción de enfermos clasificados correctamente}$$

$$\text{Especificidad} = \frac{VN}{FP + VN} = \text{proporción de sanos clasificados correctamente}$$

$$\text{VPP} = \frac{VP}{VP + FP} = \text{probabilidad que el paciente tenga la enfermedad dado que el test es positivo}$$

$VPN = \frac{VN}{VN + FN}$ = probabilidad que el paciente no tenga
la enfermedad dado que el test es negativo

Con estos conceptos presentes podemos extraer los datos que cada artículo haya declarado, existen herramientas web que pueden calcular los VP, VN, FP y FN a través de una calculadora, como lo es el programa Review Manager (Revman) de Cochrane(The Nordic Cochrane Centre, 2014), cuando estos valores no son declarados explícitamente.

Luego de tener en cuenta estos conceptos debemos de diseñar nuestra RS. Para ello existe la pauta PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)(Page et al., 2021b). Esta pauta describe 27 puntos que debemos de tener en cuenta tanto para realizar una RS y también para la lectura crítica de esta.

II. Explicación del check list de PRISMA 2020

A continuación, se presenta el check list de PRISMA 2020 para luego analizar punto por punto.

Tabla 2
Directrices PRISMA 2020 para la realización de una revisión sistemática.

Sección y tema	Artículo #	Elemento de la lista de verificación	Ubicación donde se reporta el artículo
TÍTULO			
Título	1	Identificar el informe como una revisión sistemática.	
ABSTRACTO			
Abstracto	2	Consulte la lista de verificación de PRISMA 2020 para resúmenes.	
INTRODUCCIÓN			
Razón fundamental	3	Describa la justificación de la revisión en el contexto del conocimiento existente.	
Objetivos	4	Proporcione una declaración explícita de los objetivos o preguntas que aborda la revisión.	
MÉTODOS			
Criterio de elegibilidad	5	Especifique los criterios de inclusión y exclusión para la revisión y cómo se agruparon los estudios para las síntesis.	
Fuentes de información	6	Especifique todas las bases de datos, registros, sitios web, organizaciones, listas de referencias y otras fuentes buscadas o consultadas para identificar estudios. Especifique la fecha en la que se buscó o consultó cada fuente por última vez.	
Estrategia de búsqueda	7	Presente las estrategias de búsqueda completas para todas las bases de datos, registros y sitios web, incluidos los filtros y límites utilizados.	
Proceso de selección	8	Especifique los métodos utilizados para decidir si un estudio cumplió con los criterios de inclusión de la revisión, incluido cuántos revisores examinaron cada registro y cada informe recuperado, si trabajaron de forma independiente y, si corresponde, detalles de las herramientas de automatización utilizadas en el proceso.	
Proceso de recopilación de datos	9	Especifique los métodos utilizados para recopilar datos de los informes, incluido cuántos revisores recopilaron datos de cada informe, si trabajaron de forma independiente, cualquier proceso para obtener o confirmar datos de los investigadores del estudio y, si corresponde, detalles de las herramientas de automatización utilizadas en el proceso.	
Elementos de datos	10 a	Enumere y defina todos los resultados para los cuales se buscaron datos. Especifique si se buscaron todos los resultados que fueran compatibles con cada dominio de resultados en cada estudio (por ejemplo, para todas las medidas, puntos temporales, análisis) y, en caso contrario, los métodos utilizados para decidir qué resultados recopilar.	
	10b	Enumere y defina todas las demás variables para las que se buscaron datos (por ejemplo, características de los participantes y de la intervención, fuentes de financiación). Describa cualquier suposición hecha sobre cualquier información faltante o poco clara.	
Evaluación del riesgo de sesgo del estudio.	11	Especifique los métodos utilizados para evaluar el riesgo de sesgo en los estudios incluidos, incluidos detalles de las herramientas utilizadas, cuántos revisores evaluaron cada estudio y si trabajaron de forma independiente y, si corresponde, detalles de las herramientas de automatización utilizadas en el proceso.	
Medidas de efecto	12	Especifique para cada resultado la(s) medida(s) del efecto (p.ej. índice de riesgo, diferencia de medias) utilizadas en la síntesis o presentación de los resultados.	
Métodos de síntesis	13a	Describa los procesos utilizados para decidir qué estudios fueron elegibles para cada síntesis (por ejemplo, tabular las características de la intervención del estudio y compararlas con los grupos	

Guía para la elaboración de Revisiones Sistemáticas y Metaanálisis

Sección y tema	Artículo #	Elemento de la lista de verificación	Ubicación donde se reporta el artículo
		planificados para cada síntesis (ítem #5)).	
	13b	Describa cualquier método necesario para preparar los datos para su presentación o síntesis, como el manejo de estadísticas resumidas faltantes o conversiones de datos.	
	13c	Describa cualquier método utilizado para tabular o mostrar visualmente los resultados de estudios y síntesis individuales.	
	13d	Describa cualquier método utilizado para sintetizar los resultados y proporcione una justificación de la(s) elección(es). Si se realizó un metanálisis, describa los modelos, los métodos para identificar la presencia y el alcance de la heterogeneidad estadística y los paquetes de software utilizados.	
	13e	Describa cualquier método utilizado para explorar posibles causas de heterogeneidad entre los resultados de los estudios (p. ej., análisis de subgrupos, metarregresión).	
	13f	Describa cualquier análisis de sensibilidad realizado para evaluar la solidez de los resultados sintetizados.	
Evaluación del sesgo de notificación	14	Describa cualquier método utilizado para evaluar el riesgo de sesgo debido a la falta de resultados en una síntesis (que surge de sesgos en la notificación).	
Evaluación de certeza	15	Describa cualquier método utilizado para evaluar la certeza (o confianza) en el conjunto de evidencia de un resultado.	
RESULTADOS			
Selección de estudios	16a	Describa los resultados del proceso de búsqueda y selección, desde la cantidad de registros identificados en la búsqueda hasta la cantidad de estudios incluidos en la revisión, idealmente utilizando un diagrama de flujo.	
	16b	Cite los estudios que podrían parecer cumplir con los criterios de inclusión, pero que fueron excluidos, y explique por qué fueron excluidos.	
Características del estudio	17	Cite cada estudio incluido y presente sus características.	
Riesgo de sesgo en los estudios.	18	Presentar evaluaciones del riesgo de sesgo para cada estudio incluido.	
Resultados de estudios individuales.	19	Para todos los resultados, presente, para cada estudio: (a) estadísticas resumidas para cada grupo (cuando corresponda) y (b) una estimación del efecto y su precisión (p. ej., intervalo de confianza/creíble), idealmente utilizando tablas o gráficos estructurados.	
Resultados de síntesis	20a	Para cada síntesis, resuma brevemente las características y el riesgo de sesgo entre los estudios contribuyentes.	
	20b	Presentar resultados de todas las síntesis estadísticas realizadas. Si se realizó un metanálisis, presente para cada uno la estimación resumida y su precisión (p.ej. intervalo de confianza/creíble) y medidas de heterogeneidad estadística. Si compara grupos, describa la dirección del efecto.	
	20c	Presentar los resultados de todas las investigaciones sobre posibles causas de heterogeneidad entre los resultados de los estudios.	
	20d	Presentar los resultados de todos los análisis de sensibilidad realizados para evaluar la solidez de los resultados sintetizados.	
Sesgos de información	21	Presentar evaluaciones del riesgo de sesgo debido a resultados faltantes (que surgen de sesgos en la presentación de informes) para cada síntesis evaluada.	
Certeza de la evidencia	22	Presentar evaluaciones de certeza (o confianza) en el conjunto de evidencia para cada resultado evaluado.	
DISCUSIÓN			
Discusión	23a	Proporcione una interpretación general de los resultados en el contexto de otras pruebas.	
	23b	Discuta cualquier limitación de la evidencia incluida en la revisión.	
	23c	Discuta cualquier limitación de los procesos de revisión utilizados.	
	23d	Discutir las implicaciones de los resultados para la práctica, las políticas y las investigaciones futuras.	
OTRA INFORMACIÓN			
Registro y protocolo	24a	Proporcione información de registro para la revisión, incluido el nombre y el número de registro, o indique que la revisión no fue registrada.	

Sección y tema	Artículo #	Elemento de la lista de verificación	Ubicación donde se reporta el artículo
	24b	Indique dónde se puede acceder al protocolo de revisión o indique que no se preparó un protocolo.	
	24c	Describa y explique cualquier modificación de la información proporcionada en el registro o en el protocolo.	
Apoyo	25	Describa las fuentes de apoyo financiero o no financiero para la revisión y el papel de los financiadores o patrocinadores en la revisión.	
Conflicto de intereses	26	Declare cualquier interés en competencia de los autores de la revisión.	
Disponibilidad de datos, códigos y otros materiales.	27	Indique cuáles de los siguientes están disponibles públicamente y dónde se pueden encontrar: formularios modelo de recopilación de datos; datos extraídos de los estudios incluidos; datos utilizados para todos los análisis; código analítico; cualquier otro material utilizado en la revisión.	

A continuación, explicaremos con más detalle cada punto del check list PRISMA 2020 para la elaboración de una revisión sistemática(Page et al., 2021b)

Punto 1. Título de la revisión. (PICoR):

Título debe de ser informativo integrando todos los componentes de la estrategia PICOR o PICOS, identificando si es una Revisión Sistemática o Revisión Sistemática con Metanálisis.

La estrategia PICOR o PICOS deriva de los siguientes conceptos:

P: Población o muestra

I: intervención

C: Comparador (si aplica)

Or: Objetivo respuesta (Outcome)

S: Diseño de estudio (En este caso Revisión Sistemática)

Revisemos algunos ejemplos si contienen o no la estrategia PICOR.

Ejemplo 1:

Título: Análisis de sensibilidad de la Tomografía Computada en pacientes adultos para la detección de cáncer pulmonar. Una revisión sistemática.

En este ejemplo la población viene siendo adultos con cáncer pulmonar, la intervención es la Tomografía Computada y la respuesta que se busca o el Outcome es la sensibilidad de la Tomografía Computada. el título comienza con “**análisis de sensibilidad**”, y no es muy recomendable este comienzo y se prefiere empezar diciendo: “**determinación de la precisión diagnóstica...**” , o “**eficacia de la Tomografía Computada ...**”. Ambos conceptos encierran lo que es sensibilidad, especificidad, VVP, VPN y otras pruebas de precisión.

Ejemplo 2:

Se quiere estudiar cual es la sensibilidad de la ecografía y la radiografía para la detección de neumotórax y se quiere determinar cuál es la que tienen mejor efectividad.

Título sugerente 1: Análisis de efectividad de la ecografía de tórax comparada con la radiografía de tórax para la detección de neumotórax en pacientes adultos. Una Revisión Sistemática con metaanálisis.

La población sería adultos con neumotórax, pero quizás en este caso es importante mencionar más de la condición del paciente y mencionar si este paciente es ambulatorio, de urgencia o si esta entubado etc. La intervención es la ecografía y la radiografía comparando ambas herramientas diagnósticas, la comparación ya está mencionada y el Outcome es la efectividad para la detección de neumotórax.

Título sugerente 2: Precisión de la ecografía y radiografía de tórax para la detección de neumotórax en paciente de urgencia. Una Revisión Sistemática con Metanálisis.

En este título si se consideró indicar la condición del paciente, ya que son pacientes con neumotórax en servicios de urgencia.

Título sugerente 3: Análisis de la sensibilidad y especificidad del ultrasonido y la radiografía de Tórax en la detección de neumotórax en pacientes adultos y pediátricos. Una Revisión Sistemática.

En cualquiera de los casos debe de contener la estrategia PICO_R, variando la muestra o población de estudio, variando el index test o la comparación de una técnica con otra.

Punto 2. Abstracto o resumen

Paralelamente a la RS se debe de realizar un resumen de esta. Para ello PRISMA 2020 también tiene un formato que se muestra a continuación.

Tabla 3.

Lista de verificación PRISMA para resúmenes estructurados.

Sección	Ítem	Descripción esperada
Título y objetivo	1	El título debe indicar explícitamente que se trata de una revisión sistemática (y si corresponde, un metaanálisis) de estudios de pruebas diagnósticas.
	2	Formule la pregunta de investigación, incluyendo la población estudiada, el test índice, y la condición o desenlace objetivo.
Métodos	3	Describa los criterios utilizados para seleccionar los estudios incluidos (criterios de elegibilidad).
	4	Especifique las bases de datos consultadas y la fecha en que se realizó la búsqueda.
	5	Explique qué método se usó para evaluar el riesgo de sesgo y la aplicabilidad de los estudios seleccionados.
	A1	Indique cómo se sintetizaron los resultados (ej. método estadístico o narrativo).
Resultados	6	Informe el número total de estudios incluidos, los participantes evaluados y las principales características clínicas o metodológicas.
	7	Reporte los resultados de sensibilidad y especificidad, incluyendo variabilidad, intervalos de confianza y, si aplica, los resultados del metaanálisis.
Discusión	9	Resuma las fortalezas y limitaciones metodológicas de la evidencia incluida.
	10	Presente una interpretación general de los hallazgos y su relevancia para la práctica clínica.
Otros datos	11	Indique la fuente de financiamiento o apoyo para la realización de la revisión.
	12	Señale si la revisión fue registrada previamente en alguna plataforma (ej. PROSPERO), incluyendo el número de registro.

Nota: Tabla adaptada del checklist de PRISMA 2020 para resúmenes estructurados (Page et al., 2021b)

Cada punto lo abordaremos con más detalle en la declaración extendida de PRISMA 2020. Esta lista de comprobación sirve para estructurar un resumen el cual debe de estar al inicio del escrito de RS.

Punto 3. Introducción.

Implícito en la introducción deben de ir los puntos 3 y 4

Punto 3. Describa la justificación de la revisión en el contexto del conocimiento existente.

Describir el “estado del arte”. Esto significa citar la bibliografía correspondiente a lo que se sabe hoy de la problemática en estudio, redactando él porque es importante la investigación, quienes y como se han visto beneficiados con la técnica, como se diagnostica la enfermedad que se está estudiando y cuáles han sido los resultados reportados en la actualidad. Debe de explicitar claramente cuál es la problemática de estudio, ¿los resultados de los estudios no son claros?, ¿difieren uno del otro?, ¿no se sabe con claridad cual es la precisión diagnóstica?. Etc..

Punto 4. Objetivo-pregunta de investigación.

Proporcione una declaración explícita de los objetivos o preguntas que aborda la revisión. debe de estar claramente el objetivo y la pregunta de investigación bien estructurada. Debe de tener coherencia con el título y a su vez con el objetivo general. En la introducción puede ir la pregunta de investigación o el objetivo, sin embargo en el escrito debe de estar claramente definido por separado, listo para colocarlo donde y cuando se lo pidan las editoriales o algún comité metodológico.

Con respecto a la pregunta de investigación:

Pregunta de investigación estructurada (estrategia PICO)

Toda revisión sistemática (RS) responde una pregunta clínica, esta a su vez obedece cierta estrategia que debería incluir la pregunta, como la población a estudiar, la Intervención (index test) la comparación (si es que se va a comprar una técnica con otra o una muestra con otra) y la respuesta o Outcome. Esta estrategia es la llamada “estrategia PICO” al igual que en el título.

P= población o muestra

I= Intervención (index test)

Co= comparador

R= Respuesta o Outcome

Ejemplo1:

Una pregunta sencilla sin utilizar estrategia PICO: ¿Cuál es la sensibilidad de la Tomografía para detectar cáncer pulmonar? (falta población)

Transformémosla en estrategia PICO: ¿Cuál es la sensibilidad de la tomografía computada en pacientes adultos para detectar cáncer pulmonar?

P= Pacientes adultos

I= tomografía computada

Co= no aplica

R= cáncer pulmonar

Ejemplo 2:

Sin estructurar: cual es la más sensible ¿la ecografía o la radiografía para detectar neumotórax?

Según estrategia PICO: ¿cuál es la sensibilidad de la ecografía de tórax comparada con la radiografía de tórax para detectar neumotórax en pacientes adultos?

P= paciente adulto

I=Ecografía (index test)

Co= Radiografía

R= neumotórax

Ejemplo 3:

Sin estructurar: ¿cuál es la eficacia de la Inteligencia Artificial para detectar nódulos pulmonares?

Estrategia PICO: ¿Cuál es el **algoritmo** de mayor precisión de inteligencia artificial para detectar nódulos pulmonares en imágenes de Tomografía Computada en pacientes adultos?

P=Paciente adulto

I= Algoritmo de inteligencia artificial

Co=No aplica

R= precisión para detectar nódulos pulmonares

Este es un ejemplo que a veces sucede, el index test está implícito en otro título ya que la inteligencia artificial depende varios parámetros como el algoritmo de reconocimiento o la base de datos en que se alimentó este algoritmo etc. Por ello debemos precisar cuál es el index test para realizar una búsqueda correcta de artículos a incluir en nuestra revisión. Quizás decir pacientes adultos es un espectro muy amplio, podríamos especificar la muestra según qué es lo que estamos buscando. Pacientes sospechosos de cáncer con factores de riesgo. Esto se deben de tener muy claro a la hora de formular la pregunta ya que de ello dependerán las palabras claves para la búsqueda de información.

Con respecto a los objetivos:

Objetivos (general y específicos) (PICO_R)

El objetivo general, la pregunta y el título deben de tener coherencia completa y seguir estrategia PICO_R.

¿Cuál es la sensibilidad de la tomografía computada en pacientes adultos para detectar cáncer pulmonar?

Título: análisis de sensibilidad de la Tomografía Computada en pacientes adultos para la detección de cáncer pulmonar. Una revisión sistemática.

Solo cambian algunos verbos de la pregunta de investigación.

Objetivo general: determinar la sensibilidad de Tomografía Computada en pacientes adultos para la detección de cáncer pulmonar.

Los objetivos específicos (objetivos secundarios) obedecen a como se va lograr el objetivo general

Siguiendo el Ejemplo 1:

- Describir los parámetros de adquisición de la Tomografía Computada en la detección de cáncer pulmonar
- identificar las características morfológicas del cáncer en Tomografía computada.
- Describir la correlación entre observadores para detectar cáncer pulmonar en Tomografía Computada.

MÉTODOS.

Hay que destacar que en esta sección solo se especifican o describen los puntos que veremos, no se debe de reportar resultados, los ejemplos acá dados y los gráficos, así como las tablas son para fines de ir entendiendo que es lo que se va a describir. Posteriormente a este apartado se analizará con más detalles la sección RESULTADOS, y es ahí donde se debe de reportar las tablas y gráficos aprendidos en esta sección.

A continuación, se presentan los contenidos y ejemplos desde el punto 5 al punto 15 que integran la parte de MÉTODOS en la declaración PRISMA 2020.

Punto 5. Especifique los criterios de inclusión y exclusión para la revisión y cómo se agruparon los estudios para las síntesis.

- Criterios de inclusión y exclusión

Antes de definir los criterios de inclusión y exclusión de artículos a nuestra revisión debemos de interiorizarnos en el tema en cuestión y analizar una cantidad suficiente de artículos que respondan las principales dudas que podríamos tener. Cabe la situación en que encontremos muy pocos artículos y que no respondan nuestra pregunta, en ese caso deberíamos realizar una investigación experimental y no una revisión sistemática. ¡¡Por otro lado, encontrar una revisión ya hecha y actualizada!! En ese caso buscar otro tema para realizar una revisión. También puede suceder que nos entregue más información y con esto incluir nuevos objetivos, nuevos criterios de inclusión y exclusión etc.

Una vez que ya nos interiorizamos en el tema podemos definir nuestros criterios.

Ejemplo 1:

¿Cuál la sensibilidad de la tomografía computada en pacientes adultos para detectar cáncer pulmonar?.

Título: análisis de sensibilidad de la Tomografía Computada en pacientes adultos para la detección de cáncer pulmonar. Una revisión sistemática.

Objetivo general:

Determinar la sensibilidad de Tomografía Computada en pacientes adultos para la detección de cáncer pulmonar.

Los objetivos específicos:

- Estudiar los parámetros de adquisición de la Tomografía Computada en la detección de cáncer pulmonar
- Analizar las características morfológicas del cáncer en Tomografía computada.
- Determinar la correlación entre observadores para detectar cáncer pulmonar en Tomografía Computada.

Criterios de inclusión:

- Pacientes adultos con factores de riesgo
- Artículos de los últimos 10 años
- Artículos que cuenten con la biopsia como Gold Estándar
- Artículos que hayan considerado mínimo 2 observadores para realizar el diagnóstico de cáncer pulmonar.

Criterios de exclusión.

- Estudios experimentales o de casos.
- Artículos que no declaren ciertas características que necesitamos para nuestros resultados.

En resumen, todo depende de nuestra pregunta de investigación para la Revisión Sistemática. Lo recomendable es leer acerca del tema en cuestión y hacerse expertos en el tema, con esto podrán formular de buena manera los criterios de inclusión y exclusión.

Punto 6. Fuentes de información.

Especifique todas las bases de datos, registros, sitios web, organizaciones, listas de referencias y otras fuentes buscadas o consultadas para identificar estudios. Especifique la fecha en la que se buscó o consultó cada fuente por última vez. Debe de indicar todos los sitios web o las bases de datos consultadas en la búsqueda como: Pubmed, Elsevier, Cochrane, Google académico, Scielo, etc.

Punto 7. Estrategia de búsqueda

Presente las estrategias de búsqueda completas para todas las bases de datos, registros y sitios web, incluidos los filtros y límites utilizados. **(metabuscaador y palabras claves (MESH) es obligatorio utilizar un gestor bibliográfico (ej. ZOTERO, Mendeley etc.).**

Debemos detallar qué buscadores utilizaremos y las palabras que incluiremos, tanto en español como en inglés y tratando de abarcar la mayoría de los artículos que se encuentra disponible en la web. Declarando la estrategia de búsqueda con las palabras correctas podremos asegurar que la revisión sistemática se encuentra libre del sesgo de selección de artículos o por lo menos minimizar este error, el cual nos garantice que los resultados finales de nuestra revisión sean confiables y coherentes a nuestra pregunta de investigación. Es recomendable hacer una base de datos Excel donde en las columnas estén los sitios web, palabras claves, términos Mesh, u otros términos que haya utilizado para realizar la búsqueda, además del número de

resultados arrojados por cada consulta en los sitios mencionados. Revisaremos los principales buscadores web de evidencia, y la utilización de términos Mesh que son las palabras claves que debes de utilizar para buscar los artículos potencialmente útiles para dar respuesta a la pregunta y objetivo propuesto a través de dos videos que dejo a continuación:

<https://youtu.be/uxEzCqPPeLY>

<https://youtu.be/KZczXKEJozE>

En resumen, el diseño o estrategia de búsqueda se trata de declarar donde se buscó (pubmed, Scielo, Elsevier, Cochrane, etc.), y como se buscó (términos Mesh, que son las palabras que se colocaron en el buscador para encontrar los artículos), además de esto se debe de declarar que se buscó en español y en inglés.

Luego de la búsqueda tendrás la recopilación total de artículos que posiblemente entren en el análisis de la Revisión Sistemática.

Punto 8. Proceso de selección.

Especifique los métodos utilizados para decidir si un estudio cumplió con los criterios de inclusión de la revisión, incluido cuántos revisores examinaron cada registro y cada informe recuperado, si trabajaron de forma independiente y, si corresponde, detalles de las herramientas de automatización utilizadas en el proceso.

La recogida de los artículos es recomendado que se realice por dos investigadores independientes entre sí, ya que se minimiza el sesgo de información. Luego se revisa en conjunto y se eliminan los duplicados.

Para poder describir todo el trabajo que se realizó y poder declarar todos los artículos que se excluyeron se debe de crear un diagrama de cómo se llegó a los artículos finales. Este diagrama es la recomendación PRISMA 2020 existen 4. La primera es diagrama de actualización de RS (son RS de RS), que incluye búsqueda de base de datos, registros y otras fuentes, la segunda también es actualización de RS que incluye

base de datos y registros solamente. La tercera es de nuevas RS que incluye búsqueda de base de datos, registros y otras fuentes y la cuarta también es para nuevas RS que incluye búsqueda de base de datos y registros solamente.

Los diagramas que se presentan a continuación son extraídos de referencia según Page et al., 2021. PRISMA 2020, BMJ 372:n71. Licencia CC BY 4.0.

Figura 1.

Diagrama Para RS de RS que incluyen búsqueda de base de datos, registros y otras fuentes o recursos.

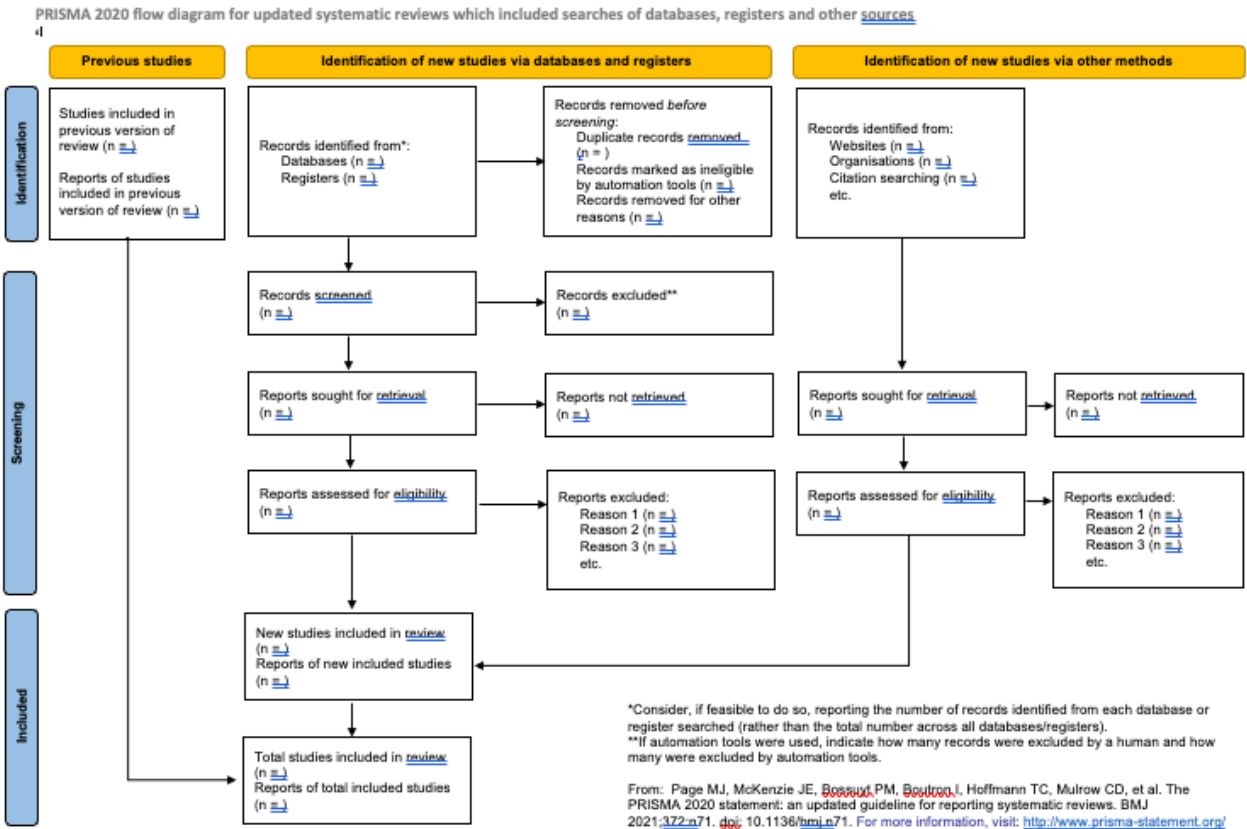
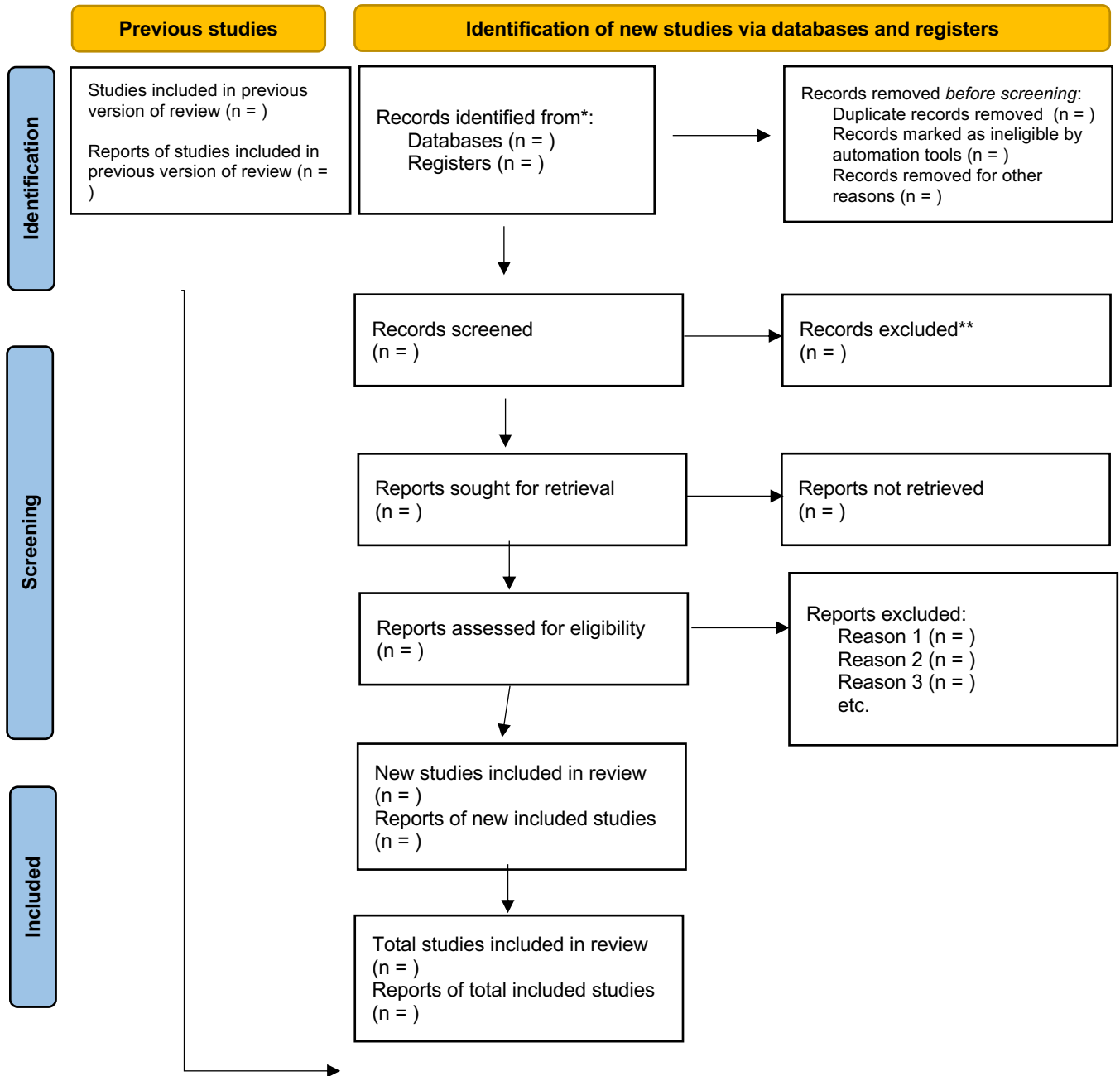


Figura 2.

Diagrama para RS de RS que incluyen búsqueda de base de datos y registros.



*Consider, if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/registers).

**If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.

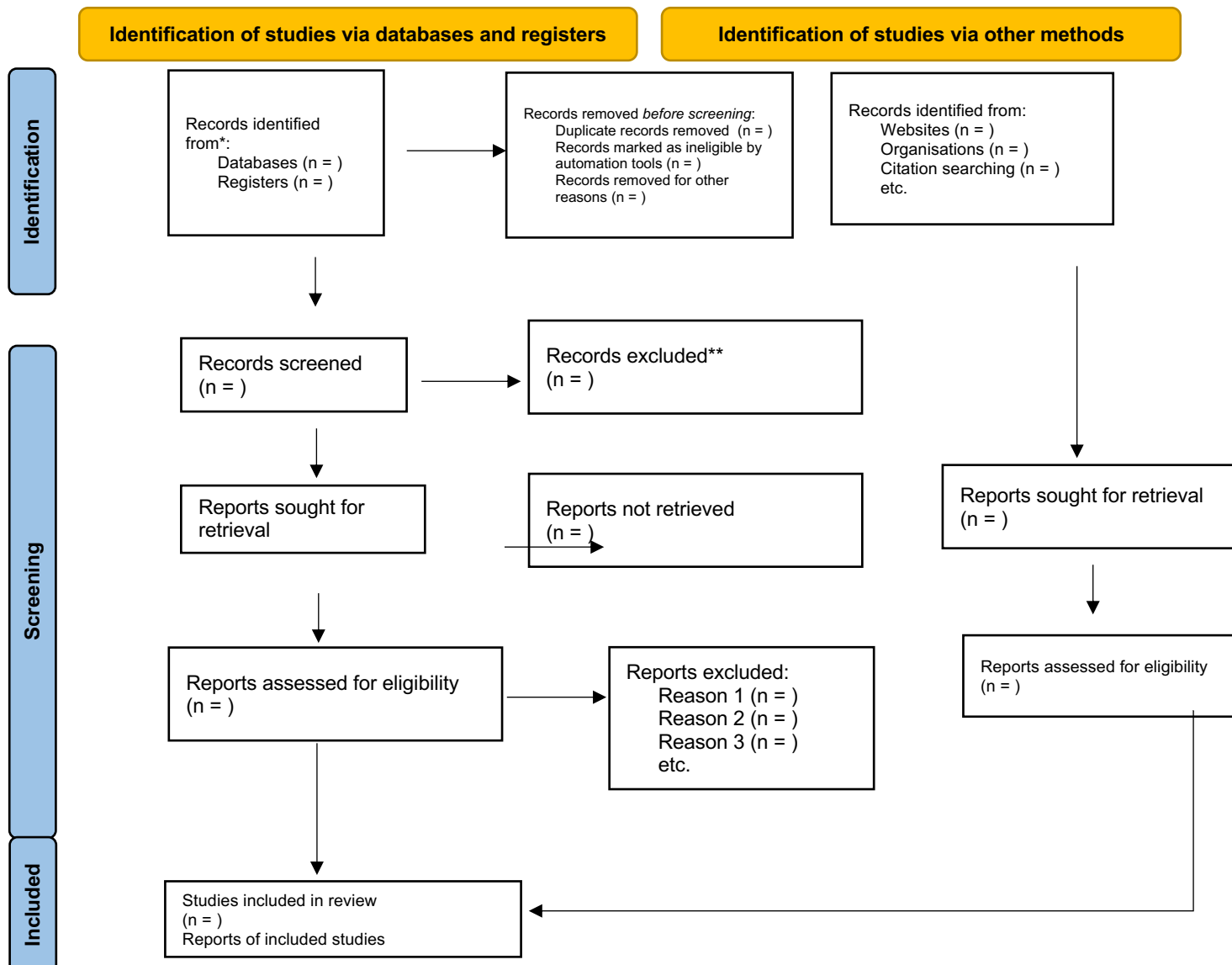
*Consider, if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/registers).

**If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71

Figura 3.

Diagrama para RS nuevas que incluyen búsqueda de base de datos, registros y otras fuentes o recursos.



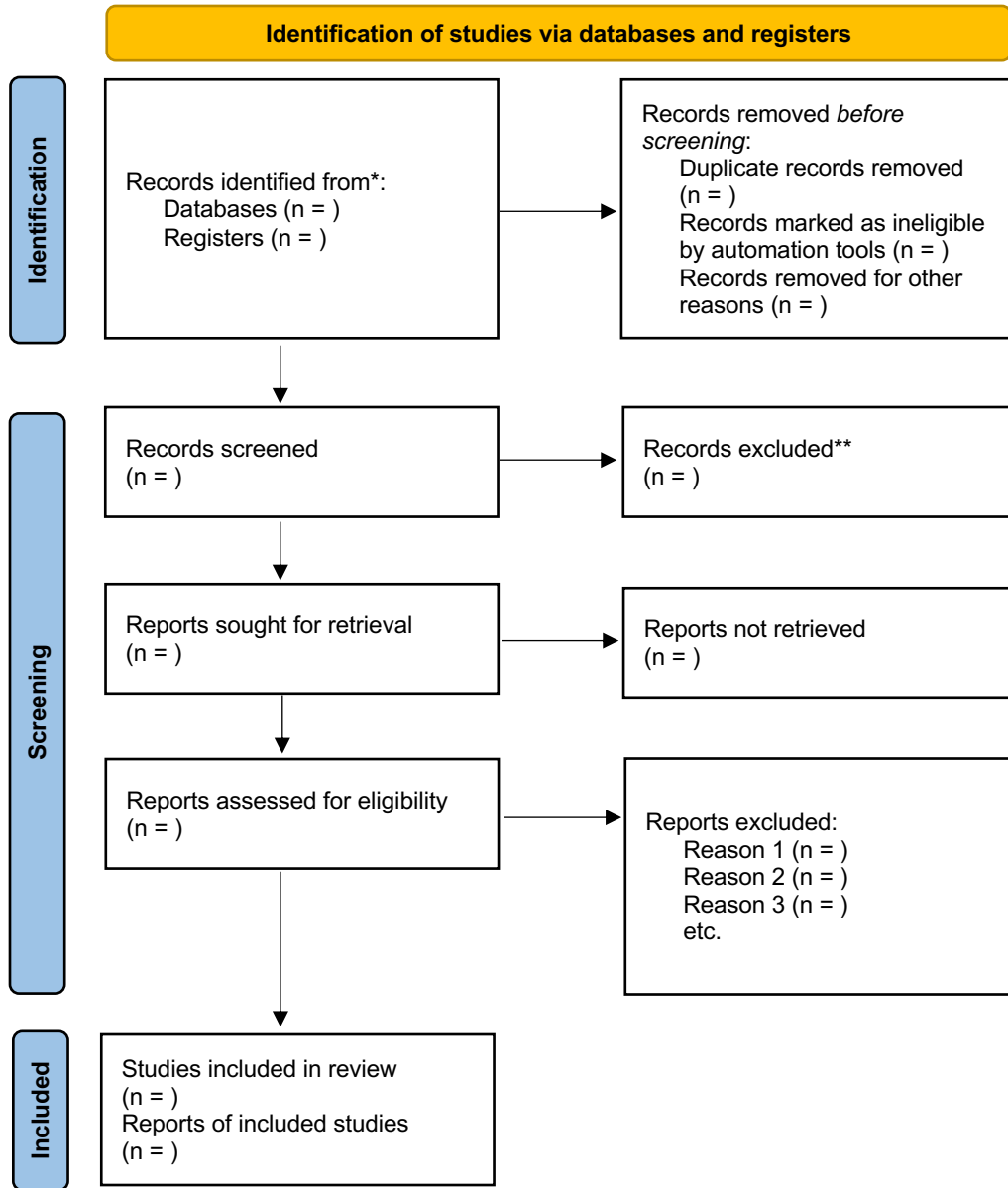
*Consider, if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/registers).

**If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. <http://www.prisma-statement.org/>

Figura 4.

Diagrama para RS nuevas que incluyen búsqueda de base de datos y registros.
(este diagrama es el que más frecuentemente se observa en nuevas RS)



*Consider, if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/registers).

**If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

Se pueden hacer estos diagramas de manera online en el siguiente enlace:
<https://www.eshackathon.org/software/PRISMA2020.html>

Punto 9. Proceso de recopilación de datos.

Especifique los métodos utilizados para recopilar datos de los informes, incluido cuántos revisores recopilaron datos de cada informe, si trabajaron de forma independiente, cualquier proceso para obtener o confirmar datos de los investigadores del estudio y, si corresponde, detalles de las herramientas de automatización utilizadas en el proceso.

Debe de indicar cuantos investigadores realizaron la extracción de datos y que metodología se ocupó (como lo hicieron). Describa si se creó una hoja Excel con las variables de estudio y fueron llenando ese formulario con los artículos seleccionados que responderán los objetivos

Considere lo siguiente:

- Desarrolle un formulario de extracción de datos que incluya variables de interés específicas.
- Extraiga los datos de manera sistemática y estandarizada de los estudios incluidos utilizando el formulario.
- Registre información sobre el diseño del estudio, características de los participantes, intervenciones, resultados y medidas de resultados, entre otros (apuntado a los objetivos específicos). Si va a realizar un metaanálisis registre los VP, VN, FP y FN.

Punto 10a y 10b. elementos de los datos.

Enumere y defina todos los resultados para los cuales se buscaron datos. Especifique si se buscaron todos los resultados que fueran compatibles con cada dominio de resultados en cada estudio (por ejemplo, para todas las medidas, puntos temporales, análisis) y, en caso contrario, los métodos utilizados para decidir qué resultados recopilar.

Enumere y defina todas las demás variables para las que se buscaron datos (por ejemplo, características de los participantes y de la intervención, fuentes de financiación). Describa cualquier suposición hecha sobre cualquier información faltante o poco clara.

Ejemplo de cómo se extraen los datos de un artículo seleccionado de una RS:

Estudio:

Título: "Evaluación de la Tomografía Computarizada de Tórax en el Diagnóstico de Neumonía en Adultos Jóvenes"

Autores (nombres ficticios):

Juan Pérez, María González, Carlos Ramírez, et al. (2020)

Características del Estudio:

Tipo de Estudio: Estudio prospectivo de cohortes.

Periodo de Estudio: Enero 2018 - Diciembre 2019.

Centro: Hospital Universitario X.

Métodos:

Participantes: Se reclutaron 150 adultos jóvenes (18-40 años) con síntomas respiratorios agudos.

Intervención: Todos los participantes se sometieron a una tomografía computarizada de tórax (TCT).

Resultado: El diagnóstico de neumonía se confirmó mediante criterios clínicos y radiológicos.

Resultados:

Sensibilidad de la TCT: 85% (IC 95%: 75% - 92%).

Especificidad de la TCT: 90% (IC 95%: 80% - 95%).

Valor Predictivo Positivo (VPP): 80%.

Valor Predictivo Negativo (VPN): 92%.

Exactitud Diagnóstica (AUC-ROC): 0.88 (IC 95%: 0.82 - 0.94).

Notas Adicionales:

Se encontraron tres falsos negativos y dos falsos positivos en la TC.

Se excluyeron los casos de neumonía atípica confirmada por pruebas específicas (por ejemplo, PCR para *Mycoplasma pneumoniae*).

Observaciones del Revisor:

La muestra del estudio parece representativa de la población objetivo.

La definición del estándar de referencia para el diagnóstico de neumonía parece apropiada.

La TC mostró una buena sensibilidad y especificidad en este estudio, pero se observaron algunos casos de resultados falsos.

Punto 11. Evaluación del riesgo del sesgo.

Especifique los métodos utilizados para evaluar el riesgo de sesgo en los estudios incluidos, incluidos detalles de las herramientas utilizadas, cuántos revisores evaluaron cada estudio y si trabajaron de forma independiente y, si corresponde, detalles de las herramientas de automatización utilizadas en el proceso.

Después de haber recopilado todos los artículos y haber excluido los artículos que no aportarán al metaanálisis o resultados de la revisión sistemática se quedarán solo unos pocos artículos. (según diagrama PRISMA).

Estos artículos son los que responderán la pregunta y objetivo de esta revisión.

A estos papers se le debe de aplicar una prueba que mida de alguna manera las falencias, debilidades o sesgos que tienen, pero de ninguna manera es una instancia

para excluir artículos, eso ya se realizó y dejó estipulado en el diagrama antes expuesto. Esta prueba es QUADAS 2 (Whiting et al., 2011) y se evalúa según sesgos y aplicabilidad que se demuestran a continuación.

Antes de aplicar QUADAS 2 a los artículos seleccionados es altamente recomendable utilizar una pauta que evalúe la calidad de reporte de cada estudio incluido, esta herramienta es STARD (Standards for Reporting Diagnostic Accuracy) (Bossuyt et al., 2015).

STARD es una guía con 30 ítems diseñada para evaluar la calidad del reporte de estudios de precisión diagnóstica (no su calidad metodológica ni su nivel de evidencia). A continuación se presenta cómo se usa paso a paso, con ejemplos, para que se pueda aplicar en una revisión sistemática.

STARD (Standards for Reporting Diagnostic Accuracy Studies)

Los 30 ítems que señala STARD para la evaluación de la calidad de reporte de cada estudio lo puede revisar en el siguiente enlace <https://www.equator-network.org/wp-content/uploads/2015/03/STARD-2015-checklist.pdf>

Ejemplo simplificado de uso práctico con 10 ítems clave:

Supongamos que estamos leyendo un estudio sobre el uso de resonancia magnética para detectar desgarros del manguito rotador.

Ítem STARD	Pregunta de evaluación	¿Se reporta?	Observaciones
1. Título	¿El título menciona que es un estudio de precisión diagnóstica?	✓	“Accuracy of MRI for rotator cuff tears”
6. Participantes	¿Describe cómo se seleccionaron los pacientes?	✗	No se menciona si fue consecutivo o aleatorio
9. Prueba índice	¿Describe cómo se aplicó la RM?	✓	Describe parámetros técnicos y operador
10. Referencia estándar	¿Describe bien el gold standard (artroscopia)?	✓	Artroscopia mencionada como referencia
13. Cegamiento	¿Se cegó al evaluador de la RM respecto al gold standard?	✗	No se indica

Ítem STARD	Pregunta de evaluación	¿Se reporta?	Observaciones
15. Análisis estadístico	¿Describe cómo se calcularon sensibilidad y especificidad?	✓	Incluye fórmulas y software
19. Flujo de pacientes	¿Presenta un diagrama de flujo de pacientes incluidos y excluidos?	✗	Solo describe el total de pacientes
23. Resultados	¿Se reportan sensibilidad, especificidad, VPP, VPN con IC95%?	✓	Se reportan con intervalos de confianza
25. Limitaciones	¿Discute las limitaciones del estudio?	✓	Señala que no todos fueron sometidos al gold standard
30. Financiamiento	¿Se informa la fuente de financiamiento o conflictos de interés?	✗	No se menciona

STARD ayuda a saber si un artículo reporta suficiente información como para ser evaluado con QUADAS-2. Si falta información, STARD lo evidencia.

Imaginemos que se aplica QUADAS-2 a un estudio y se observa que no queda claro si hubo cegamiento del evaluador. Si revisamos STARD y el ítem 13 (ciego a la referencia estándar) no está reportado, se puede justificar que hay riesgo de sesgo incierto por información no reportada.

Esto fortalece la transparencia y trazabilidad de la evaluación en la revisión sistemática.

Evaluación de sesgos y aplicabilidad según QUADAS2

QUADAS 2 Son figuras que evalúan el riesgo de sesgos de “cada artículo seleccionado” para responder los objetivos del trabajo.

Paso 1: Evaluación del título y abstract de los paper seleccionados: el artículo a incluir debe de tener la población de estudio, el index test (la prueba a evaluar) y el test de referencia, si no tiene estos componentes no vale la pena incluirlo en la revisión final.

Una vez que filtramos los artículos según abstract, título, criterios de inclusión y exclusión por parte de los investigadores nos quedaremos con los artículos finales que se incluirán en los resultados de la revisión sistemática.

Paso 2: Diagrama de flujo: revisar los flujos, reclutamiento y condición, esto debe de ser reportado por el investigador en cada artículo, si no aparece el investigador de la revisión sistemática debe de realizarlo.

Antes de analizar los pasos 3 y 4 de QUADAS2 revisaremos los sesgos más frecuentes y fases de un estudio diagnóstico, ya que los estudios en fase 1 quizás no son recomendados para incluirlos en los resultados finales ya que tienen sesgos que los veremos a continuación a través de unos ejemplos:

Sesgo de selección de pacientes: una prueba óptima sobre cierta condición es la que se emplea en condiciones que el paciente pueda confundirse. Dicho de otro modo, que los pacientes a estudio sean sospechosos de la condición o potencialmente susceptibles. No obstante, las pruebas de diagnóstico se dividen en 4 fases entonces dependiendo de lo que estamos estudiando es la selección de paciente. Las fases son las siguientes:

Fase I: es cuando se somete a estudio una prueba diagnóstico nuevo y no sabemos si es capaz de discriminar entre sanos y enfermos. En este caso, se realizan estudios de casos y controles y se evalúa si existen diferencias entre ellos. Estos estudios reportan valores de sensibilidad, especificidad y exactitud de la prueba. En esta fase los resultados atienden la pregunta: ¿los enfermos difieren de los sanos? Estos estudios tienen grandes sesgos de selección de pacientes ya que tenemos por un lado pacientes muy enfermos y por otros pacientes muy sanos, entonces faltan los pacientes sospechosos, pero como no sabemos nada de la prueba y no gastar tantos recursos económicos y humanos se hace esta primera fase de test diagnósticos de

casos y controles. A pesar de que son muy importantes para tener los primeros datos estadísticos de la nueva prueba o test diagnóstico este tipo de artículos deben de ser excluidos de RS por lo mencionado anteriormente.

Fase II: en esta etapa al igual que en la fase I se realizan en casos y controles por ello está en una situación “ideal” ya que no tenemos casos sospechosos por lo tanto la prueba no puede responder si realmente es capaz de discriminar cierta enfermedad. Esta etapa sirve para realizar puntos de corte, y es el primer paso para un estudio diagnóstico. En esta fase se atiende la pregunta ¿cuál es la probabilidad de un resultado de la prueba dado el estado del sujeto? En esta fase al igual que la primera tienen sesgo de selección de pacientes y tampoco debe de ser incluida en RS.

Fase III: esta fase atiende la pregunta: en pacientes susceptibles de tener la enfermedad, ¿los resultados de la prueba discriminan aquellos con y sin la enfermedad?

Para dar respuesta a esta pregunta necesitamos saber la “realidad” o tener una prueba de referencia llamada “Gold Estándar”.

Esta es la fase que mejor representa los estudios diagnósticos en radiología, ya que la **fase IV** se trata del pronóstico del paciente, esto quiere decir, si el paciente se sometió a la prueba ¿presenta mejor pronóstico que el que no se sometió a la prueba diagnóstica?

Este tipo de estudios es el que debemos de ser capaces de identificar en la literatura e incluirlos en una RS (criterio de inclusión).

Con respecto al estándar de referencia o Gold Estándar se pueden presentar errores como la prueba de referencia no es la más adecuada, a esto se le llama **sesgo de estándar de referencia, que muchas veces puede ser la biopsia y no en todas las investigaciones se cuenta con ella.**

A modo de ejemplo:

Se estudia la capacidad de la Tomo síntesis para detectar y clasificar correctamente nódulos mamarios tanto malignos como benignos. El estudio reclutó 50 pacientes, dentro de las imágenes se cuenta con todo tipo de nódulos, tanto malignos como benignos e indeterminados (fase III estudio diagnóstico) y la correspondiente biopsia como Gold Estándar. Estas imágenes ¿quiénes las evaluarán?, radiólogos especialistas en radiología de mama? ¿Con cuanta experiencia? ¿Serán ciegos a la biopsia cuando analicen las imágenes?, cuantos radiólogos evaluarán las imágenes? ¿Se incluyeron las mamas densas? ¿El radiólogo 1 concuerda con el radiólogo 2? (variabilidad inter-observador) Cada radiólogo observa siempre lo mismo? (variabilidad intra-observador).

Un artículo que evalúa este tipo de procedimientos debería responder estas preguntas, si no es así podría haber eventuales “sesgos” en el estudio.

Otro ejemplo es la comparación entre dos técnicas: supóngase que se quiere saber que precisión tiene la ecografía de tórax comparada con la radiografía en la detección de neumotórax. ¿importa el tiempo entre ambos estudios?

El artículo por revisar debería ser visto con ojo crítico por parte del investigador que realiza una revisión sistemática y él determinar la importancia y el nivel del artículo a incluir según el tipo de pregunta y los objetivos de la revisión sistemática. Detectar los sesgos es una tarea fundamental y QUADAS2 lo que pretende es identificar estas debilidades en los estudios

Una vez que ya hemos revisado cuales son los sesgos más comunes de encontrar y revisar los tipos de estudios diagnósticos, revisaremos los pasos 3 y 4 de QUADAS2.

Paso 3 y 4 riesgo de sesgos y aplicabilidad:

Selección de pacientes:

Riesgo de sesgo: ¿Podría la selección de pacientes haber introducido sesgos?

1. ¿Se enroló una muestra consecutiva o aleatoria de pacientes?
2. ¿Se evitó un diseño de casos y controles?
3. ¿Se evitaron exclusiones inapropiadas? Por ejemplo, se excluyeron pacientes de difícil diagnóstico o que incluyen pacientes con la enfermedad confirmada y no con la sospecha de la enfermedad.

Aplicabilidad: ¿Hay preocupación de que la aplicación o la interpretación de la prueba que está siendo evaluada no coincidan con la pregunta de la revisión?

Puede ser que los pacientes incluidos en el estudio fueron diferentes a los contemplados en la pregunta de investigación respecto a su condición, sus características demográficas etc. (ej: los tumores más grandes son más fáciles de identificar por lo tanto la sensibilidad de la prueba será sobrestimada)

Prueba índice:

Riesgo de sesgo: ¿podría la realización o la interpretación de la prueba índice haber introducido sesgos?

- 1.- fueron interpretados los resultados de la prueba índice sin conocimiento de los resultados de la de referencia?
- 2.- si se utilizó un umbral para definir positividad o la negatividad de la prueba índice ¿fue especificado previamente?

Aplicabilidad: ¿hay preocupación que la conducción o interpretación de la prueba índice no coincida con la pregunta de investigación?

Si los métodos de la prueba índice varían con respecto a lo especificado en la pregunta de revisión puede existir duda acerca de su aplicabilidad.

Prueba de referencia:

Riesgo de sesgo: ¿Podría la interpretación de la prueba de referencia haber introducido sesgos?

1.- ¿Es probable que la prueba de referencia valore correctamente la patología o condición del paciente?

2.- ¿Fueron interpretados los resultados de la prueba de referencia sin conocimiento de la prueba índice?

Aplicabilidad: ¿Hay preocupación de que la condición diana, clasificada como tal a través de la prueba de referencia difiera de la población a la cual estaba referida la pregunta?

Ejemplo: en la definición de infección urinaria la prueba de referencia se basa en urocultivo. Sin embargo, el umbral por encima del cual el resultado se considera positivo puede variar.

Flujo y tiempos:

Riesgo de sesgo: ¿podría el flujo de paciente haber introducido sesgos?

1.- ¿hubo intervalo apropiado entre la prueba índice y la prueba de referencia? Hay veces en que se necesita que la prueba índice y la de referencia debe de recolectarse en el mismo momento. A la inversa una prueba de referencia que implica un tiempo de seguimiento puede requerir un período mínimo para evaluar si la condición está presente.

2.- ¿fue aplicada a todos los individuos la misma prueba de referencia? (sesgo de verificación)

3.- ¿fueron incluidos todos los pacientes en el análisis? Puede haber sesgos si el número de pacientes incluidos en el estudio difieren de los pacientes incluidos en los resultados reportados.

QUADAS2 elaboró una tabla para determinar los riesgos de sesgos y su aplicabilidad dependiendo si es alta, baja o inespecífica, y es la que se debe de reportar en el artículo de revisión. También se recomienda acompañar esta tabla de un gráfico que a

continuación se presenta. *Cabe destacar que este checklist de QUADAS2 no asigna puntaje ni tampoco es un criterio para eliminar artículos, en esta fase solo se “transparenta” la información y se declara la calidad de artículos que responderán los objetivos de la revisión sistemática. La fase de eliminación de artículos se debe realizar leyendo abstract y título según criterios de inclusión y exclusión.*

Pauta QUADAS 2

Tabla 4.

Análisis de riesgo de sesgo y problema de aplicabilidad QUADAS-2

Estudio	Riesgo de sesgo: Selección de pacientes	Riesgo de sesgo: Test índice	Riesgo de sesgo: Estándar de referencia	Riesgo de sesgo: Flujo y tiempo	Problemas de aplicabilidad: Selección de pacientes	Problemas de aplicabilidad: Test índice	Problemas de aplicabilidad: Estándar de referencia
Estudio 1	😊😊😊😊	😊😊😊😊	😊😊😊😊	😊😊😊😊	😞😞😞😞	😊😊😊😊	😊😊😊😊
Estudio 2	😊😊😊😊	😊😊😊😊	😊😊😊😊	😊😊😊😊	😞😞😞😞	😊😊😊😊	😊😊😊😊
Estudio 3	😊😊😊😊	😊😊😊😊	😊😊😊😊	😊😊😊😊	😞😞😞😞	😊😊😊😊	😊😊😊😊
Estudio 4	😊😊😊😊	😊😊😊😊	😊😊😊😊	😊😊😊😊	😞😞😞😞	😊😊😊😊	😊😊😊😊
Estudio 5	😊😊😊😊	😊😊😊😊	😊😊😊😊	😊😊😊😊	😞😞😞😞	😊😊😊😊	😊😊😊😊
Estudio 6	😊😊😊😊	😊😊😊😊	😊😊😊😊	😊😊😊😊	😞😞😞😞	😊😊😊😊	😊😊😊😊

Nota: Las figuras en verde están representando bajo nivel de sesgo, en rojo alto nivel de sesgo y en amarillo incertidumbre (poco claro). Esquema adaptado según criterios metodológicos de QUADAS-2 (Whiting et al., 2011).

Este es un ejemplo de nuestra tabla sugerida para resumir los resultados de su evaluación QUADAS-2. Esto se puede adaptar a su revisión copiando y pegando los símbolos apropiados y usándolos para completar la tabla (en esta etapa se evalúan los artículos no se excluyen).

Finalmente se debe de reportar un gráfico generalizado, ya no por estudios, si no es una representación total de todos los estudios incluidos, tal cual se muestra a continuación.

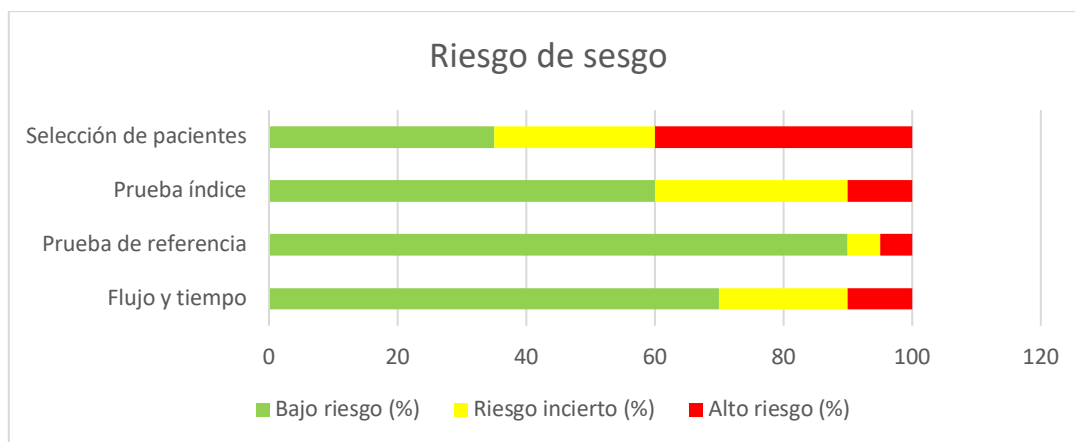
El siguiente gráfico es un ejemplo tomando solo el riesgo de sesgo, pero se debe de hacer también para problemas de aplicabilidad. El gráfico se creó adaptando el porcentaje de figuras que tienen bajo nivel de sesgo, alto nivel de sesgo y figuras que representan incertidumbre, a través de tabla y gráfico en Excel.

Se recomienda revisar el siguiente video: <https://youtu.be/LsclmYA1H3Q>

Tabla 5.
Riesgo de sesgo generalizada

Dominio	Bajo riesgo (%)	Riesgo incierto (%)	Alto riesgo (%)
Flujo y tiempo	70	20	10
Prueba de referencia	90	5	5
Prueba índice	60	30	10
Selección de pacientes	35	25	40

Figura 5.
Gráfico según la tabla de riesgo de sesgo



Nota: Solo este gráfico se reporta, la tabla anterior es para describir de cómo se puede generar este gráfico de colores (faltaría complementar con problemas de aplicabilidad en otro gráfico).
Gráfico adaptado de referencia (Whiting et al., 2011).

Tabla 6.
Ejemplo para evaluar el riesgo de sesgo

Dominio	1. Selección de participantes	2. Prueba de índice	3. Estándar de referencia	4. Flujo y sincronización
Preguntas y criterios de señalización	P1: " ¿Se inscribió una muestra consecutiva o aleatoria de participantes? "	P1: " ¿Se interpretaron los resultados de la prueba índice sin conocimiento de los resultados del estándar de referencia? "	P1: " ¿Es probable que el estándar de referencia clasifique correctamente la condición objetivo? "	P1: " ¿Hubo un intervalo adecuado entre la prueba índice y el estándar de referencia? "
	Sí, si el estudio informa sobre una selección	Sí, si el estudio informa que los resultados de la prueba índice se	Sí, si el estándar de referencia define correctamente la presencia/ausencia de	Sí, si el intervalo entre la prueba índice y el estándar

	<p>consecutiva o aleatoria de participantes.</p> <p>No – si el estudio informa sobre otra forma de selección de participantes.</p> <p>Poco claro: si el estudio no informa cómo se inscribieron los participantes.</p> <p>P2: "¿El estudio evitó exclusiones inapropiadas?"</p> <p>Sí, si las definiciones de los criterios de exclusión son apropiadas (es decir, cirugía o tratamiento previo para el CHC; personas con colangiocarcinoma) y se informan todas las exclusiones.</p> <p>No, si los criterios de exclusión son inapropiados y no se informan las exclusiones.</p> <p>Poco claro: si el estudio no informa las causas de las exclusiones.</p>	<p>interpretaron sin el conocimiento de los resultados del estándar de referencia.</p> <p>No, si el estudio informa que los resultados de la prueba índice se interpretaron con los resultados del estándar de referencia.</p> <p>Poco claro: si el estudio no proporciona información sobre el cegamiento de los resultados de la prueba índice y el estándar de referencia.</p> <p>P2: "¿Se definieron claramente los criterios de positividad?"</p> <p>Sí, si el estudio informa claramente los criterios de positividad (es decir, para agentes de células de Kupffer cuando se detecta hiperrealce en la fase arterial y un lavado tardío (≥ 60 segundos). Para agentes de células de Kupffer cuando aparecen abundantes vasos tumorales en forma de cesta. Se detectan formas ramificadas similares o irregulares desde la periferia hasta el centro de la lesión, y tinción tumoral densa en la fase vascular temprana y lavado rápido en la fase vascular tardía, y defecto de Kupffer completo).</p> <p>No – si el estudio no reporta los criterios de positividad.</p>	<p>CHC, como patología del hígado explantado en una cohorte de trasplante).</p> <p>No, si se utilizaron otras pruebas de referencia además de la patología del hígado explantado, como la histología de la muestra reseca o la biopsia de lesión focal.</p> <p>P2: "¿Se interpretaron los resultados del estándar de referencia sin el conocimiento de los resultados de la prueba índice?"</p> <p>Sí, si el estudio informa que los resultados del estándar de referencia se interpretaron sin el conocimiento de los resultados de la prueba índice.</p> <p>No, si el estudio informa que los resultados del estándar de referencia se interpretaron con conocimiento de los resultados de la prueba índice.</p> <p>Poco claro: si el estudio no proporciona información sobre el cegamiento de los resultados del estándar de referencia y la prueba índice.</p>	<p>de referencia fue inferior a 3 meses.</p> <p>No – si el intervalo fue igual o superior a 3 meses.</p> <p>Poco claro: si el estudio no informa el intervalo entre la prueba índice y el estándar de referencia.</p> <p>P2: "¿Recibieron todos los participantes el mismo estándar de referencia?"</p> <p>Sí, si el estudio tiene solo 1 estándar de referencia para todos los participantes.</p> <p>No, si el estudio tiene > 1 estándar de referencia.</p> <p>Poco claro: si la información del estudio sobre el uso del estándar de referencia no está clara.</p> <p>Pregunta 3: "¿Se incluyeron todos los participantes en el análisis y se analizaron según el principio de intención de diagnosticar (los resultados no evaluables se consideraron falsos)?"</p> <p>Sí, si todos los participantes inscritos estuvieran incluidos en el análisis.</p> <p>No, si algún participante fue excluido del análisis por algún motivo.</p> <p>Poco claro: si la exclusión de los participantes del</p>
--	--	--	---	---

				<p>análisis no está clara.</p> <p>P4: " <i>¿Se incluyeron y analizaron los participantes con resultados no evaluables de la prueba índice de acuerdo con el principio de intención de diagnosticar (los resultados no evaluables se consideraron falsos)?</i> "</p> <p>Sí, si los participantes con resultados no evaluables se incluyeron y analizaron según el principio de intención de diagnosticar.</p> <p>No: si los participantes con resultados no evaluables no fueron incluidos ni analizados según el principio de intención de diagnosticar.</p>
Riesgo de sesgo	<p><i>¿Podría la selección de participantes haber introducido sesgos?</i></p> <p>Si respondemos "sí" a todas las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "bajo".</p> <p>Si respondemos "no" a ≥ 1 de las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "alto".</p> <p>Si respondemos "poco claro" a todas las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "poco claro".</p> <p>Si respondemos "poco claro" a ≥ 1 de las preguntas de señalización y a las restantes nuestra respuesta es "sí",</p>	<p><i>¿Podría la realización o interpretación de la prueba índice haber introducido sesgo?</i></p> <p>Si respondemos "sí" a todas las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "bajo".</p> <p>Si respondemos "no" a ≥ 1 de las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "alto".</p> <p>Si respondemos "poco claro" a todas las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "poco claro".</p> <p>Si respondemos "poco claro" a ≥ 1 de las preguntas de señalización y a las restantes nuestra respuesta es "sí", entonces</p>	<p><i>¿Podría el estándar de referencia, su conducta o su interpretación haber introducido sesgos?</i></p> <p>Si respondemos "sí" a todas las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "bajo".</p> <p>Si respondemos "no" a ≥ 1 de las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "alto".</p> <p>Si respondemos "poco claro" a todas las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "poco claro".</p> <p>Si respondemos "poco claro" a ≥ 1 de las preguntas de señalización y a las restantes nuestra respuesta es "sí", entonces</p>	<p><i>¿Podría el flujo de participantes haber introducido sesgos?</i></p> <p>Si respondemos "sí" a todas las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "bajo".</p> <p>Si respondemos "no" a ≥ 1 de las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "alto".</p> <p>Si respondemos "poco claro" a todas las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "poco claro".</p> <p>Si respondemos "poco claro" a todas las preguntas de señalización, entonces consideramos que el riesgo de sesgo es "poco claro".</p> <p>Si respondemos "poco claro" a ≥ 1 de las preguntas de señalización y a las restantes nuestra respuesta es "sí", entonces</p>

	entonces consideramos que el riesgo de sesgo es "poco claro".	consideramos que el riesgo de sesgo es "poco claro".	consideramos que el riesgo de sesgo es "poco claro".	riesgo de sesgo es "poco claro". Si respondemos "poco claro" a ≥ 1 de las preguntas de señalización y a las restantes nuestra respuesta es "sí", entonces consideramos que el riesgo de sesgo es "poco claro".
Preocupaciones sobre la aplicabilidad	<p><i>¿Existe la preocupación de que los participantes incluidos y el entorno no coincidan con la pregunta de revisión?</i></p> <p>Preocupación baja: los participantes incluidos en la revisión representan a los participantes en quienes se utiliza la prueba en la práctica clínica (es decir, modalidad de imágenes de segunda línea en personas con sospecha de lesión hepática).</p> <p>Preocupación alta: los participantes incluidos en la revisión difieren de los participantes en quienes se utiliza la prueba en la práctica clínica.</p>	<p><i>¿Existe la preocupación de que la prueba índice, su realización o interpretación difieran de la pregunta de revisión?</i></p> <p>Preocupación baja: la prueba índice, su realización o su interpretación no difieren de la forma en que se utiliza en la práctica clínica.</p> <p>Gran preocupación: la prueba índice, su realización o su interpretación difieren de la forma en que se utiliza en la práctica clínica.</p>	<p><i>¿Existe la preocupación de que la condición objetivo definida por el estándar de referencia no coincida con la pregunta?</i></p> <p>Gran preocupación: la definición de la condición objetivo según lo definido por el estándar de referencia no coincide con la pregunta (es decir, la patología del hígado explantado es factible sólo en el caso de un trasplante de hígado; la historia natural y el pronóstico del CHC detectado en el hígado explantado podrían ser diferente).</p> <p>Preocupación baja: la definición de la afección objetivo según la define el estándar de referencia coincide con la pregunta, p. ej., tomografía computarizada o resonancia magnética para todos los participantes incluidos.</p>	

Punto 12- medidas de efecto.

Especifique para cada resultado la(s) medida(s) del efecto (p.ej. índice de riesgo, diferencia de medias) utilizadas en la síntesis o presentación de los resultados.

Debe de reportar cual fue la medida que respondieron los objetivos del estudio, esto debe de estar claro desde el principio de la investigación ya que lo debe de tener presente en los criterios de inclusión. Estas medidas pueden ser las siguientes:

ODD RATIO (OR): Son medidas de riesgo el cual indica la probabilidad del grupo de expuestos a la condición desarrollen la enfermedad vs los no expuestos. Generalmente se usa esta medida de efecto en estudios donde se quiere saber si existe relación de una enfermedad con alguna exposición. En RS con meta-análisis (RSM) se suele observar esta medida en estudios de tratamiento o de eficacia de algún fármaco.

Proporciones: Se observa reportada en RSM de estimación de prevalencia. Recordemos que la prevalencia es la proporción de individuos que tienen cierta condición dividido por el total de la población. Entonces se puede hacer una RSM teniendo muy en cuenta que la prevalencia de alguna enfermedad está condicionada por el tiempo y muchas veces por el lugar o país donde se realiza el estudio individual. es muy importante definir en los criterios de elegibilidad (inclusión y exclusión) estos parámetros de tiempo y lugar de búsqueda de la evidencia. Existe una guía para evaluar la calidad en estudios de prevalencia(Migliavaca et al., 2020) .

Precisión diagnóstica: En las RS con o sin Metanálisis de precisión diagnóstica se reportan una pareja de resultados que es la sensibilidad y la especificidad, que a diferencia de las RS de tratamiento reportan solo uno que es el OR. El meta-análisis estará condicionado por la homogeneidad de los artículos, es decir, que hayan utilizado el mismo index test, el mismo estándar de referencia y en pacientes similares (Zamora Romero et al., 2009).

Además de reportar la sensibilidad y la especificidad se reporta la curva ROC resumen (sROC) a través de una medida global que es el área bajo la curva (AUC) el cual resume el valor de la curva en un solo número que va del 0 a 1 donde el AUC más cercano a 1 es mejor para la prueba mientras cercano a 0.5 la prueba no tendría capacidad de discriminación entre sanos y enfermos.

Diferencia de medias: En este tipo de RS se evalúa el impacto de una intervención o exposición en una variable continua, por ejemplo la diferencia de peso después de una intervención, la presión arterial, las dosis de radiación, etc.. por lo tanto se debe de definir claramente que es lo que se va a estimar o reportar en la RS en los criterios de elegibilidad, describiendo claramente la medida o unidad que se pretende resumir.

Punto 13 a,b,c,d,f. Métodos de síntesis

13a. Describa los procesos utilizados para decidir qué estudios fueron elegibles para cada síntesis (por ejemplo, tabular las características de la intervención del estudio y compararlas con los grupos planificados para cada síntesis (ítem #5)).

13b. Describa cualquier método necesario para preparar los datos para su presentación o síntesis, como el manejo de estadísticas resumidas faltantes o conversiones de datos.

Describa cualquier método utilizado para tabular o mostrar visualmente los resultados de estudios y síntesis individuales.

13c. Describa cualquier método utilizado para sintetizar los resultados y proporcione una justificación de la(s) elección(es). Si se realizó un Metaanálisis, describa los modelos, los métodos para identificar la presencia y el alcance de la heterogeneidad estadística y los paquetes de software utilizados.

13d. Describa cualquier método utilizado para explorar posibles causas de heterogeneidad entre los resultados de los estudios (p. ej., análisis de subgrupos, metarregresión).

13c. Describa cualquier análisis de sensibilidad realizado para evaluar la solidez de los resultados sintetizados.

Tal cual como se indica, debe de describir todos los métodos en los cuales se van a reportar los resultados. Como, por ejemplo, describa si se van a realizar tablas resúmenes para reportar las características de los artículos seleccionados para

responder los objetivos (número de pacientes, tipo de estudio, conclusiones de los autores, número de eventos reportados, número de pacientes con cierta condición etc.). En el caso de realizar Metaanálisis debe de reportar que tipo de estadística es la que se va a realizar (efectos fijos o aleatorios) y que pruebas. En el caso de pruebas de diagnóstico debe de indicar que se realizó Forest Plot (gráfico de bosque) para resumir la sensibilidad y la especificidad, además de curvas sROC para determinar el área bajo la curva de la prueba. Cuando se realiza el gráfico de Forest Plot, este aparte de reportar la sensibilidad y especificidad de la prueba también reporta índice de heterogeneidad I^2 . El paquete estadístico por excelencia para pruebas de diagnóstico es STATA (*Statistical software for data science* | *Stata*, s. f.). En resumen, todo lo que va a ser declarado en los resultados debe de estar indicado previamente paso a paso, resultado por resultado y como es que se elaboró.

Punto 14. Evaluación sesgo de notificación.

Describa cualquier método utilizado para evaluar el riesgo de sesgo debido a la falta de resultados en una síntesis (que surge de sesgos en la notificación).

En una revisión sistemática de pruebas de diagnóstico, un ejemplo de sesgo de notificación podría surgir cuando los investigadores conocen previamente los resultados de la prueba de referencia antes de realizar la prueba diagnóstica bajo estudio. Este conocimiento previo podría influir en la interpretación de los resultados de la prueba diagnóstica, ya sea de manera consciente o inconsciente, lo que lleva a una sobreestimación de la precisión diagnóstica.

Por ejemplo, supongamos que en un estudio se realiza una prueba de diagnóstico de imagen (por ejemplo, resonancia magnética) para detectar la presencia de una lesión específica en el hombro. Sin embargo, los investigadores conocen los resultados de una prueba de referencia más precisa (por ejemplo, biopsia) antes de realizar la resonancia magnética. Si los resultados de la biopsia se le comunican al radiólogo que

interpreta las imágenes de resonancia magnética, esto podría sesgar la interpretación de las imágenes, ya que el radiólogo podría ser consciente de la ubicación exacta de la lesión y buscar características específicas en la resonancia magnética que coincidan con la biopsia, lo que podría aumentar artificialmente la sensibilidad y especificidad de la prueba diagnóstica en comparación con una situación en la que el radiólogo esté ciego a los resultados de la biopsia.

Este tipo de sesgo de notificación puede distorsionar los resultados de la revisión sistemática al inflar artificialmente la precisión diagnóstica de la prueba bajo estudio, lo que destaca la importancia de realizar la interpretación de las pruebas de diagnóstico de manera ciega y prospectiva para minimizar el riesgo de sesgo de notificación.

Este tipo de sesgo también es abordado en QUADAS2, por lo tanto, se puede declarar que se utilizó QUADAS2 para evaluar riesgo de sesgo.

Punto 15. Evaluación de la certeza.

Describa cualquier método utilizado para evaluar la certeza (o confianza) en el conjunto de evidencia de un resultado.

Mencionamos en un comienzo que las RS (con o sin Metaanálisis) están en el contexto de la toma de decisiones, por lo tanto se hace necesario una evaluación de la evidencia en este contexto. Para clasificar la evidencia es que nació GRADE en sus siglas en inglés (Grades of Recommendation, Assessment, Development, and Evaluation). Instituciones como la colaboración Cochrane la han utilizado para definir recomendaciones en torno a la evidencia estudiada.

Las recomendaciones se realizan según 4 niveles de GRADE (Balslem et al., 2011).

Tabla 7.

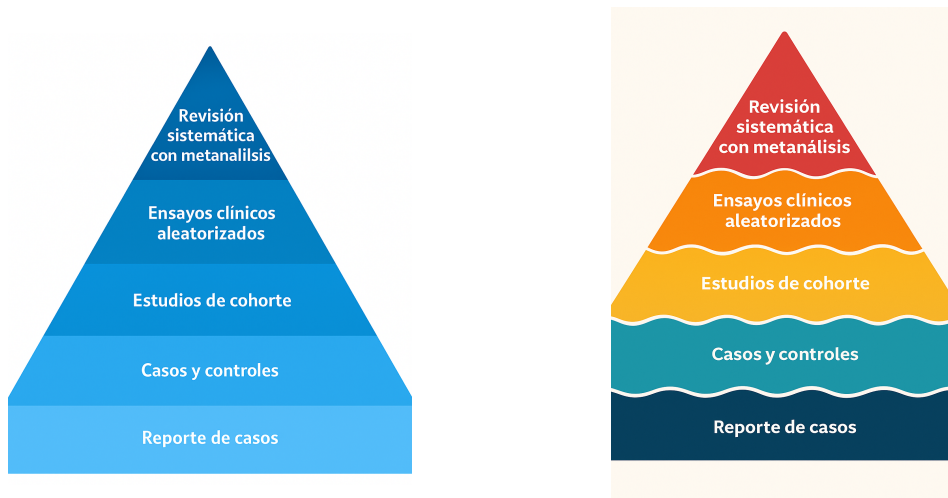
Nivel de certeza y significado

Nivel de la certeza	significado
Alto	Estamos muy seguros que el verdadero efecto se acerca al estimado
Moderado	Tenemos una confianza moderada en el efecto, el efecto verdadero puede que este cerca de la estimación del efecto, pero hay probabilidades que sea sustancialmente diferente.
Bajo	Nuestra confianza en el efecto es limitada, el verdadero efecto puede ser sustancialmente distinto al estimado
Muy bajo	Tenemos muy poca confianza en la estimación del efecto, el verdadero efecto puede ser muy distinto al estimado

Ya hemos hablado de los diseños de estudios y un Ensayo Clínico Controlado Aleatorizado (ECA) están la cima de la pirámide de la evidencia, mientras que los estudios observacionales están por debajo. Sin embargo en un nuevo concepto de la evidencia podemos observar en la siguiente figura que en determinados casos se pueden dar fluctuaciones de la evidencia según su diseño.

Figura 6.

Pirámide de la evidencia.



Nota : imágenes adaptadas de Murad et al., (2016)

GRADE propuso que la calidad de los estudios no está solo determinada según la pirámide de evidencia, sino que es un conjunto de conceptos que puede desafiar el concepto de pirámide de la evidencia (Murad et al., 2016).

La figura de la izquierda representa la pirámide de la evidencia tradicional. En la figura de la derecha se observan los niveles de la evidencia de forma “ondulada”, donde pueden existir ciertos estudios evaluados de mejor manera que los que están por arriba, además se observan las RS/metaanálisis en la cima de la evidencia. No obstante, toda la evidencia debe de ser analizada de forma rigurosa.

Entonces, cómo funciona la evaluación GRADE: la evaluación depende de ciertos conceptos que hacen subir o bajar el nivel de evidencia existente (Balslem et al., 2011). Hay que considerar que QUADAS 2 solo evalúa una parte de la certeza de la evidencia que es el riesgo de sesgo, por lo tanto se deben de evaluar más partes, y eso hace GRADE.

A continuación, explicaremos el paso a paso con GRADE.

Paso a paso de la evaluación de la calidad de la evidencia según GRADE

Punto de partida: Tipo de estudio

Tabla 7
Evaluación de GRADE según tipo de estudio.

Tipo de estudio	Nivel inicial en GRADE	Notas
Ensayo clínico aleatorizado (ECA)	⊕⊕⊕⊕ (Alta)	Por tener control y aleatorización
Estudio de cohorte (prospectivo o retrospectivo)	⊕⊕○○ (Baja)	Observacional: puede mejorar si está bien hecho
Estudio de casos y controles	⊕⊕○○ (Baja)	Observacional, más propenso a sesgos
Serie de casos	⊕○○○ (Muy baja)	No tiene grupo comparador
Reporte de caso	⊕○○○ (Muy baja)	Menor nivel de evidencia
Estudios ecológicos / de tendencia	⊕○○○ (Muy baja)	No permiten atribuir causa-efecto
Opinión de expertos / consensos	⊕○○○ (Muy baja)	Sin datos empíricos, pero puede ser útil como contexto

Paso 2: conceptos que hacen bajar el nivel de certeza.

Tabla 8
5 conceptos que reducen la certeza

Concepto que reduce certeza	¿Qué evalúa?	¿Cuándo se baja un nivel?	¿Cuándo se bajan dos niveles?
Riesgo de sesgo	Calidad metodológica de los estudios (evaluado por QUADAS-2 o Cochrane)	Cuando varios estudios tienen sesgo leve a moderado	Cuando la mayoría tiene alto riesgo de sesgo grave
Inconsistencia	Variabilidad entre estudios (medido por I ² o diferencias visuales)	Heterogeneidad moderada (I ² > 30%) o resultados dispares	Heterogeneidad severa (I ² > 75%) y sin explicación
Indirectidad	Diferencias entre lo evaluado y lo que realmente interesa (paciente, intervención, test o desenlace)	Población o test no exactamente iguales al foco clínico	Múltiples diferencias importantes o indirectas
Imprecisión	Incertidumbre del resultado (ancho del intervalo de confianza)	Intervalos amplios que cruzan umbrales clínicos	Intervalos muy amplios o con muestra insuficiente
Sesgo de publicación	Falta de estudios negativos o tendencia a publicar solo efectos positivos	Funnel plot algo asimétrico o sospecha moderada	Funnel plot claramente asimétrico o evidencia directa

Nota: Cada uno de los 5 factores puede hacerte bajar 1 o hasta 2 niveles, según la gravedad del problema.

Paso 3: Conceptos que hacen subir la certeza

Tabla 9
Conceptos que mejoran la certeza

Factor que aumenta	Mejora	Ejemplo práctico
Efecto grande	+1 o +2	OR > 5 o AUC > 0.95
Gradiente dosis-respuesta	+1	Más dosis = mejor diagnóstico
Confusores plausibles	+1 (efecto sería más fuerte sin ellos)	Todos los sesgos van contra el efecto observado

Nota: Si se tiene estudios observacionales pero cumplen esto, podrías subir de “baja” a “moderada” o incluso a “alta”.

Paso 4: ¿Cómo se ve el resultado final?

Tabla 10
Reporte gráfico de GRADE

Nivel final	Símbolo GRADE
Alta	⊕⊕⊕⊕
Moderada	⊕⊕⊕○
Baja	⊕⊕○○
Muy baja	⊕○○○

Ejemplo práctico

- Partes con un estudio observacional → ⊕⊕○○ (**baja**)
- No hay sesgo, no hay inconsistencia → no baja.
- El efecto es muy grande (AUC = 0.96) → subes +1 → ⊕⊕⊕○ → **Moderada**

EJEMPLO PASO A PASO: APLICACIÓN DE GRADE EN UNA REVISIÓN SISTEMÁTICA DE PRECISIÓN DIAGNÓSTICA

Pregunta de investigación: ¿Qué tan precisa es la resonancia magnética (RM) para diagnosticar lesiones del manguito rotador en pacientes con hombro doloroso?

Resultado evaluado: Sensibilidad

Se incluyen 6 estudios primarios en la revisión sistemática.

PASO 1: Tipo de estudio (nivel inicial de certeza)

Estudio	Diseño
1	Estudio de cohorte
2	Estudio de cohorte
3	Estudio de casos y controles
4	Ensayo clínico aleatorizado
5	Estudio de cohorte
6	Estudio de cohorte

Como la mayoría (4/6) son **observacionales**, el nivel inicial de certeza **comienza en BAJA** (⊕⊕○○).

PASO 2: Evaluar si hay factores que hagan bajar la certeza

Dominio GRADE	Evaluación	¿Se baja el nivel?
Riesgo de sesgo	QUADAS-2 muestra que 4 estudios tienen bajo riesgo, 2 moderado	✗ No baja
Inconsistencia	Sensibilidades entre 0.85 y 0.92, I ² = 25% (baja heterogeneidad)	✗ No baja
Imprecisión	IC agrupado de sensibilidad = 0.89 (IC 95%: 0.86–0.92)	✗ No baja
Indirectez	Todos los estudios usan RM 1.5T, misma población (clínica)	✗ No baja
Sesgo de publicación	Funnel plot simétrico	✗ No baja

Resultado parcial: Se mantiene en **BAJA certeza** (⊕⊕○○)

Paso 3: Evaluar si hay factores que hagan subir la certeza

Factor de aumento	¿Aplica?	Justificación
Efecto muy grande	✓ Sí	Sensibilidad agrupada = 0.89, con IC estrecho
Relación dosis-respuesta	✗ No	No evaluada
Todos los sesgos subestiman	✗ No	No se documenta

Se sube **1 nivel** por efecto muy grande → pasa a **MODERADA certeza** (⊕⊕⊕○)

Paso 4: RESULTADO FINAL GRADE: Sensibilidad de la RM

Resultado	Certeza final	Justificación GRADE
Sensibilidad	⊕⊕⊕○ Moderada	Estudios observacionales con bajo sesgo, IC estrecho y efecto fuerte

Esta evaluación se presenta en una tabla GRADE y se menciona en el apartado de resultados de la revisión sistemática.

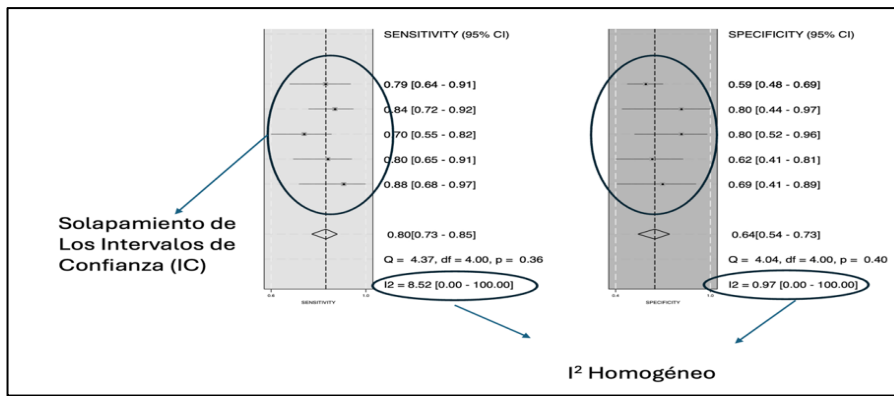
GRADE no se basa en cuán "impresionante" es el resultado, sino en **qué tan confiable es**. Gracias a eso, se puede tener **confianza moderada** en que la sensibilidad real de la RM para detectar lesiones del manguito rotador ronda el 89%.

El ejemplo que se planteó se basa en un resultado que es la sensibilidad, se debe de hacer lo mismo para cada objetivo planteado, si otro objetivo era especificidad,

valores predictivos, sobrevida, efectividad etc., se debe de hacer un análisis GRADE específico para cada objetivo planteado.

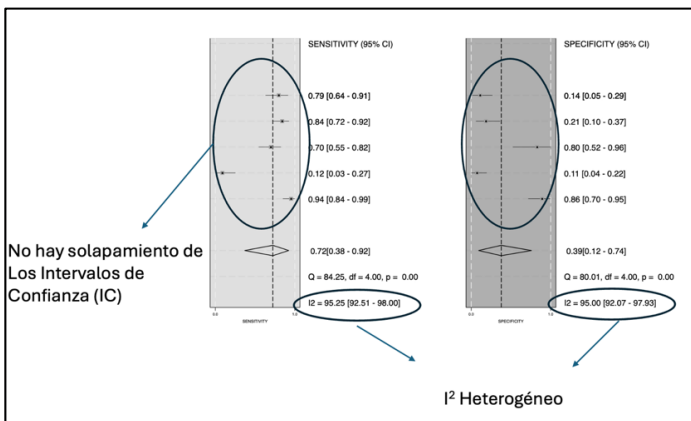
Se ha mencionado en distintas oportunidades la heterogeneidad de los estudios. A continuación, se presentan algunos ejemplos para interpretarlos correctamente.

Ejemplos de RSM con homogeneidad y heterogeneidad de sus artículos (realizado por programa STATA v18):



La homogeneidad de los artículos se puede observar en el forest plot por el solapamiento de los IC (óvalo negro), tanto en la sensibilidad como en la especificidad. En la parte estadística el I² es de 8,52 para la sensibilidad y un 0,97 para la especificidad indicando una homogeneidad consistente.

Ejemplo de heterogeneidad:



A diferencia de los IC del estudio con los artículos homogéneos, estos no se solapan entre ellos y es más evidente en los IC de la especificidad, podemos decir “cualitativamente” que existe heterogeneidad de los artículos incluidos en el metaanálisis. Estadísticamente el I^2 es del 95% tanto para la sensibilidad como la especificidad indicando una alta heterogeneidad coincidiendo con lo que vemos cualitativamente en los IC.

En este escenario no es conveniente realizar un metaanálisis de los resultados y se recomienda un análisis cualitativo.

RESULTADOS:

Esta sección depende del diseño de estudio. Si es una revisión sistemática sin Metanálisis se pueden elaborar una serie de tablas que incluyan a cada autor con los resultados más relevantes que expliquen el objetivo de estudio.

Si es un Metanálisis de pruebas diagnósticas donde se quiere reportar sensibilidad y especificidad o se quiere comparar dos técnicas se deben de tener en consideración agregar en los criterios de inclusión artículos que reporten los verdaderos positivos (VP), los verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN), ya que con estos datos se construye gráfico de forest plot y sus respectivas sensibilidades y especificidades. Otro análisis estadístico que no debe de faltar son las curvas ROC y el Área Bajo la Curva (AUC), que además de reportar el poder de discriminación de la prueba también reporta si los estudios son homogéneos y es una manera bastante más fácil de visualizar la comparación de pruebas diagnósticos.

A continuación, se presentan del punto 16 al punto 22 correspondiente a la sección de resultados.

Punto 16a: Selección de estudios:

Se debe describir cómo se seleccionaron los estudios desde la búsqueda inicial hasta los artículos incluidos en el análisis final. Idealmente esto se presenta mediante un diagrama PRISMA 2020 actualizado (ver Figuras 1,2,3 y 4).

- Ejemplo: “Se identificaron 327 artículos en la búsqueda inicial. Tras eliminar duplicados y aplicar los criterios de elegibilidad, se seleccionaron 18 estudios para la síntesis cualitativa y 12 de ellos fueron incluidos en el metaanálisis (ver Diagrama PRISMA en Figura 3).” Así debería ir la descripción luego de poner el diagrama PRISMA correspondiente a la RS que se está revisando (ver Punto 8)

Punto 16b: Estudios excluidos con justificación:

Aquí se deben citar los estudios potencialmente elegibles que fueron excluidos en etapa final, junto con el motivo de exclusión.

- Ejemplo: “Se excluyeron 3 estudios por utilizar un estándar de referencia diferente al definido en la revisión (biopsia). Además, 2 estudios fueron excluidos por no reportar valores de VP, VN, FP y FN, impidiendo el análisis de precisión diagnóstica.”

Punto 17: Características de los estudios incluidos:

Se deben presentar las características relevantes de los estudios incluidos. Esto puede mostrarse en una tabla que incluya autores, año, país, muestra, test índice, estándar de referencia, tipo de estudio, y conclusión principal.

- Ejemplo: Tabla 5 muestra los 12 estudios incluidos, de los cuales 8 corresponden a estudios de cohortes y 4 a estudios caso-control. La mayoría fue realizada en hospitales de tercer nivel y todos utilizaron biopsia como test de referencia.

Punto 18: Evaluación del riesgo de sesgo:

Este ítem debe mostrar cómo se evaluó la calidad metodológica de los estudios incluidos. Se recomienda utilizar QUADAS-2 y presentar tanto tabla como figura resumen. (tal cual la mostramos en la tabla 4 y figura 4)

- Ejemplo: “El análisis de riesgo de sesgo mediante QUADAS-2 evidenció bajo riesgo en los dominios de flujo y tiempo y estándar de referencia. Sin embargo, hubo preocupación en la selección de pacientes en 3 estudios y en la aplicabilidad del test índice en 2 estudios. Ver Figura X.”

Punto 19: Resultados de estudios individuales:

Este apartado incluye los resultados desglosados por estudio: sensibilidad, especificidad, valores predictivos, AUC, etc. Se recomienda usar tablas y gráficos estructurados.

- Ejemplo: “La sensibilidad de la tomografía computada para detectar nódulos pulmonares varió entre 0.75 y 0.92, mientras que la especificidad osciló entre 0.80 y 0.96. En Figura X se presentan los valores individuales de sensibilidad y especificidad para cada estudio incluido.”

Punto 20a: Resultados de síntesis (resumen general)

Presenta un resumen global de la síntesis: cuántos estudios, riesgo general de sesgo, características generales, etc.

- Ejemplo: “Se sintetizaron 12 estudios con un total de 2.300 pacientes. En general, los estudios presentaron bajo riesgo de sesgo y buena concordancia en el uso del test índice y el estándar de referencia.”

Punto 20b: Resultados estadísticos del metaanálisis

Este es el núcleo del análisis estadístico. Se deben presentar las estimaciones agrupadas (sensibilidad/especificidad) con sus intervalos de confianza, curva ROC resumen (sROC), área bajo la curva (AUC) y medidas de heterogeneidad (I^2).

- Ejemplo: “La sensibilidad agrupada fue de 0.88 (IC 95%: 0.85–0.91) y la especificidad agrupada de 0.91 (IC 95%: 0.88–0.94). El área bajo la curva sROC fue 0.94, indicando alta precisión. La heterogeneidad fue moderada ($I^2 = 32%$ para sensibilidad, $I^2 = 40%$ para especificidad). Ver Figura 6.”

Punto 20c: Investigaciones sobre heterogeneidad

Reportar cualquier análisis exploratorio de fuentes de heterogeneidad (subgrupos, meta regresión, etc.).

- Ejemplo: “Se realizó un análisis por subgrupo según tipo de equipo (TC multicorte vs. helicoidal) mostrando mejores resultados para los equipos multicorte ($p < 0.05$).”

Punto 20d: Análisis de sensibilidad

Se debe indicar si se excluyeron estudios con alto riesgo para comprobar la robustez de los resultados.

- Ejemplo: “Al excluir 3 estudios con alto riesgo de sesgo, los resultados globales no se modificaron significativamente, confirmando la robustez de los hallazgos.”

Punto 21: Sesgo de publicación (sesgos de información)

Reportar si se utilizó alguna técnica para detectar sesgo de publicación, como gráfico de embudo (funnel plot).

- Ejemplo: “El análisis del gráfico de embudo no mostró asimetría importante, lo que sugiere bajo riesgo de sesgo de publicación.”

Punto 22: Certeza de la evidencia (GRADE)

Se presenta la calidad global de la evidencia para cada resultado, usando la metodología GRADE (alta, moderada, baja, muy baja).

- Ejemplo: “La certeza de la evidencia fue alta para la sensibilidad y moderada para la especificidad, según criterios GRADE (riesgo de sesgo bajo, precisión adecuada y consistencia entre estudios). Ver Tabla 6.”

EJEMPLO DE REPORTE DE RESULTADOS EN REVISIONES SISTEMÁTICAS CON METANÁLISIS.

A continuación, se ejemplifica cómo debe estructurarse y redactarse el apartado de resultados en una revisión sistemática y Metaanálisis de precisión diagnóstica, siguiendo los puntos 16 al 22 de la guía PRISMA 2020. Se utiliza como ejemplo ficticio una revisión sobre la precisión diagnóstica de la resonancia magnética (RM) para detectar lesiones del manguito rotador en pacientes con hombro doloroso.

1.- Selección de estudios:

Se identificaron 356 registros en las bases de datos. Tras eliminar duplicados (n = 104) y aplicar los criterios de elegibilidad, se evaluaron 24 artículos a texto completo. Finalmente, se incluyeron 6 estudios en la síntesis cuantitativa (Metaanálisis).

Figura 1. Diagrama de flujo PRISMA 2020

(Insertar diagrama con los cuatro bloques: identificación, cribado, elegibilidad, incluidos)

2.- Estudios excluidos con justificación

De los 24 estudios evaluados a texto completo, se excluyeron 18:

- 6 por no reportar datos de sensibilidad y especificidad
- 5 por no usar RM como test índice
- 4 por no utilizar artroscopía como estándar de referencia
- 3 por duplicación de población en otros artículos

3.- Características de los estudios incluidos:

Los estudios incluidos fueron realizados entre 2010 y 2022. La mayoría fueron estudios de cohorte prospectiva (n = 4), uno fue ensayo clínico y uno estudio de casos y controles. Todos utilizaron RM 1.5T como test índice y artroscopía como estándar de referencia.

Tabla X. Características de los estudios incluidos

Estudio	País	Diseño	Muestra	Test índice	Estándar de referencia	Sensibilidad	Especificidad
A	EE.UU.	Cohorte	120	RM 1.5T	Artroscopía	0.88	0.84
B	Japón	Casos y controles	92	RM 1.5T	Artroscopía	0.90	0.80
C	Chile	Cohorte	100	RM 1.5T	Artroscopía	0.91	0.88
D	España	Ensayo clínico	70	RM 1.5T	Artroscopía	0.85	0.86
E	Brasil	Cohorte	150	RM 1.5T	Artroscopía	0.92	0.87
F	Canadá	Cohorte	110	RM 1.5T	Artroscopía	0.89	0.85

4.- Evaluación del riesgo de sesgo (QUADAS-2)

Se aplicó la herramienta QUADAS-2. Cuatro estudios fueron evaluados con bajo riesgo en todos los dominios, mientras que dos mostraron preocupación moderada en el dominio de selección de pacientes.

Figura X. Resumen gráfico de QUADAS-2 por dominio (Insertar gráfico de barras por dominio: selección de pacientes, test índice, estándar de referencia, flujo y tiempo)

Tabla X. Evaluación tabular de QUADAS-2 (Insertar tabla con color o símbolos tipo 😊 / ? / 😞 para cada dominio y cada estudio).

Agregar además que se complementó el análisis de la calidad del reporte mediante check list de STARD (disponible como anexo).

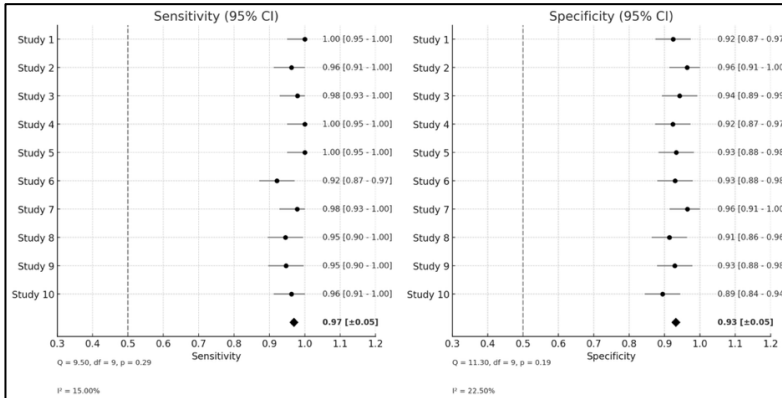
5.- Resultados individuales de los estudios:

Las sensibilidades individuales oscilaron entre 0.85 y 0.92, mientras que las especificidades variaron entre 0.80 y 0.88. Estos valores fueron graficados mediante forest plots.

Figura X. Forest Plot de Sensibilidad y Especificidad por estudio

(Insertar forest plot con IC para cada estudio y efecto agrupado como diamante)

Ejemplo:

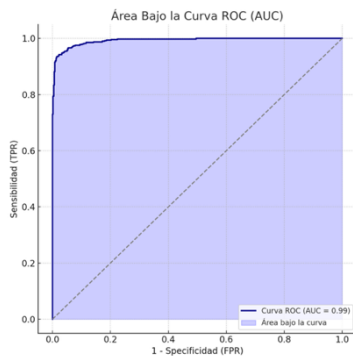


6.- Resultados de la síntesis estadística

El Metaanálisis de efectos aleatorios mostró:

- Sensibilidad agrupada: 0.89 (IC 95%: 0.86–0.92)
- Especificidad agrupada: 0.85 (IC 95%: 0.81–0.89)
- AUC (curva ROC resumen): 0.93
- Heterogeneidad: $I^2 = 22\%$ (sensibilidad), $I^2 = 28\%$ (especificidad)

Figura X. Curva ROC resumen (sROC)



7.- Análisis de heterogeneidad:

Se realizó un análisis por subgrupos según tipo de equipo (RM 1.5T convencional vs. RM de alto campo). No se observaron diferencias significativas entre subgrupos (p = 0.15)

8.- Análisis de sensibilidad:

Excluyendo los dos estudios con riesgo de sesgo moderado, los resultados se mantuvieron estables: sensibilidad = 0.90, especificidad = 0.86.

9.- Sesgo de publicación:

El gráfico de embudo (funnel plot) mostró simetría. No se evidenció sesgo de publicación.

Figura X. Funnel Plot de estudios incluidos

10.- Evaluación de la certeza de la evidencia (GRADE)

Se aplicó GRADE al resultado de sensibilidad:

- Diseño: estudios observacionales (nivel inicial = bajo)
- No se bajó el nivel por riesgo de sesgo, imprecisión ni inconsistencia
- Se subió un nivel por efecto grande y consistente

Tabla X. Certeza de la evidencia según GRADE

Resultado	Diseño inicial	Problemas	Aumento	Certeza final	Justificación
Sensibilidad	Observacional	Ninguno	+1	⊕⊕⊕○ Moderada	Efecto grande, IC estrecho

Resultado final: **Moderada certeza (⊕⊕⊕○)**

A continuación, se ejemplifica el reporte de resultados de Revisiones Sistemáticas sin Metanálisis. Se señalarán solo los pasos que se diferencian del reporte anteriormente visto.

EJEMPLO DE REPORTE DE RESULTADOS EN REVISIONES SISTEMÁTICAS SIN METANÁLISIS.

1.- Selección de estudios: Igual que el ejemplo con anterior

2.- Estudios excluidos con justificación: igual que el ejemplo anterior

3.- Características de los estudios incluidos: Agregue en la última columna **notas relevantes.**

4.- Evaluación del riesgo de sesgo (QUADAS-2): Igual que el ejemplo anterior.

5.- Resultados individuales de los estudios

Se presenta un resumen narrativo por estudio:

- **Estudio A:** Sensibilidad 0.88, especificidad 0.82, RM 1.5T convencional
- **Estudio D:** Sensibilidad 0.76, especificidad 0.74, bajo campo → menor rendimiento
- **Estudio E:** Sensibilidad 0.94, especificidad 0.87, RM 3T con secuencias avanzadas

No se agruparon estadísticamente los datos debido a diferencias clínicas y metodológicas.

6.- Resultados de la síntesis cualitativa

La mayoría de los estudios mostró sensibilidades entre 0.85 y 0.94, con especificidades entre 0.80 y 0.87. Un estudio (D) tuvo valores notablemente menores, probablemente por el uso de tecnología de campo bajo.

Tabla X. Resumen descriptivo de los hallazgos.

7.- Heterogeneidad y análisis de sensibilidad

No se realizó análisis estadístico de heterogeneidad ni análisis de sensibilidad formal. Se describe que los resultados más discordantes se asocian a diferencias en el tipo de equipo y metodología de evaluación.

8.- Sesgo de publicación

Dado que no se realizó Metanálisis, no se construyó funnel plot. Sin embargo, la revisión manual no detectó ausencia evidente de estudios negativos.

9.- Evaluación de la certeza GRADE: Igual que el ejemplo anterior.

A continuación se ejemplifican los últimos dos apartados de las directrices PRISMA 2020, que son la **discusión y otra información**, que van desde el punto 23 al 27.

DISCUSIÓN DE UNA RS:

Se discuten las implicaciones clínicas de los hallazgos, incluyendo la utilidad de las pruebas de diagnóstico en diferentes contextos clínicos y poblaciones.

Se destacan las limitaciones de la evidencia encontrada, como el riesgo de sesgo en los estudios incluidos y la variabilidad en los resultados reportados.

Se sugieren áreas para futuras investigaciones, como la comparación directa entre diferentes pruebas de diagnóstico y la evaluación de su impacto en los resultados clínicos.

Punto 23a: Interpretación general de los resultados

Ejemplo: La sensibilidad de la RM para diagnosticar lesiones del manguito rotador mostró un rendimiento alto y consistente en la mayoría de los estudios incluidos (sensibilidad entre 0.85 y 0.94). Estos resultados son comparables con estudios previos de metaanálisis en población similar, lo que refuerza su aplicabilidad clínica.

Punto 23b: Limitaciones de la evidencia

Ejemplo: Las principales limitaciones identificadas fueron:

- La mayoría de los estudios fueron observacionales.
- Variabilidad en el tipo de RM utilizada (campo bajo vs. alto campo).
- Diferencias en experiencia del evaluador y secuencias empleadas.

Estas limitaciones podrían influir levemente en la precisión real del test.

Punto 23c: Limitaciones del proceso de revisión

Ejemplo: las principales limitaciones del proceso de revisión fueron:

- No se realizó búsqueda en literatura gris.
- La revisión fue realizada por dos autores, sin tercer revisor externo.
- No se contactó a los autores para datos faltantes.

Estas limitaciones podrían haber restringido la cantidad de evidencia incluida.

Punto 23d: Implicancias para la práctica y futuras investigaciones

Ejemplo: Los hallazgos apoyan el uso de la RM como herramienta precisa para el diagnóstico de lesiones del manguito rotador. No obstante, se recomienda:

- Establecer protocolos estandarizados de imagen.
- Comparar rendimiento con ecografía y nuevas tecnologías (IA, imagen 3T).
- Incluir poblaciones más diversas y estudios multicéntricos.

OTRA INFORMACIÓN RELEVANTE SEGÚN PRISMA 2020

Punto 24a: Registro de la revisión

Ejemplo: La revisión fue registrada previamente en PROSPERO con el número de registro: CRD42023455678.

Punto 24b: Acceso al protocolo

Ejemplo: El protocolo de la revisión está disponible en línea a través de PROSPERO en el siguiente enlace:

https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42023455678

Punto 24c: Modificaciones del protocolo

Ejemplo: Durante la ejecución de la revisión, se agregó una búsqueda manual en Google Scholar, que no estaba contemplada inicialmente. No hubo otros cambios relevantes.

Punto 25: Apoyo financiero

Ejemplo: Esta revisión no recibió financiamiento externo. Fue desarrollada como parte del trabajo docente universitario del autor principal.

Punto 26: Conflicto de intereses

Ejemplo: Los autores declaran no tener ningún conflicto de intereses en relación con esta revisión.

Punto 27: Disponibilidad de datos y materiales

Ejemplo: Los siguientes materiales están disponibles previa solicitud al autor:

- Formulario de extracción de datos.
- Base de datos en Excel con sensibilidad/especificidad.
- Evaluaciones QUADAS-2.
- Tabla GRADE final.

Además, los datos analíticos (gráficos y resultados) fueron generados con STATA 18 y están archivados para revisión por pares.

CONSIDERACIONES FINALES DEL AUTOR

Aunque previamente se revisó la estructura formal de un resumen para una revisión sistemática (punto 2), es importante señalar que no se abordó de manera didáctica cómo garantizar que no se omita información relevante. Una estrategia útil para asegurar un resumen completo y claro, especialmente útil en congresos, seminarios o presentaciones orales. Consiste en redactarlo a partir de una secuencia de preguntas clave, que permitan cubrir todos los elementos esenciales del estudio.

Este enfoque facilita la comprensión del trabajo tanto para audiencias académicas como generales, y asegura que se incluyan los aspectos más relevantes.

Esquema de contenidos para realizar un resumen de una RS.

Ejemplo extraído de referencia (Rud et al., 2012):

1.- **Resumen (Abstract)** dentro del resumen considerar lo siguiente:

1.1.- **Antecedentes** (Background)

1.2.- **objetivos:**

- primarios
- Secundarios

1.3.- **Métodos de búsqueda:** donde de buscó, fecha de búsqueda, sin restricción de idioma

1.4.- **Criterios de selección:** criterios de elegibilidad, inclusión y exclusión, además indicar que la búsqueda se realizó por dos investigadores de manera independiente.

1.5.- **Recopilación de datos y análisis:** indicar que se evaluó la calidad metodológica de los artículos incluidos a través de la Evaluación de Calidad de artículos de Precisión Diagnóstica en sus siglas en inglés QUADAS2. Además, se evaluó la calidad de reporte mediante el check list de STARD. Además, reportar que se usó el modelo de efectos aleatorios para obtener el resumen de sensibilidad y especificidad de la prueba.

1.6.- **Resultados importantes:** incluir número de artículos seleccionados, número de participantes y resultados más importantes incluyendo intervalos de confianza.

1.7.- **Conclusiones de los autores.**

Resumen a través de preguntas y respuestas:

2.1.- pregunta de investigación: ejemplo: ¿Cuál es la precisión diagnóstica de la tomografía computada en la pancreatitis aguda?

2.2.- ¿Porque es importante mejorar el diagnóstico de la patología estudiada?

ejemplo: ¿porque es importante mejorar el diagnóstico de la pancreatitis?

R. La TC permite evaluar y diferenciar entre personas que necesitan operarse y las que no lo necesitan.

2.3 ¿Cuál es el objetivo de la revisión?

Ejemplo: el objetivo es evaluar la precisión diagnóstica de la TC en el diagnóstico de la pancreatitis aguda.

2.4.- ¿Qué se estudió en la revisión?

R.- se analizó la visualización del páncreas con medio de contraste endovenoso y si podía mejorar con Tomografía de alta resolución, además se evaluó la calidad de imagen con protocolos de bajas de dosis de radiación.

2.5.- ¿Cuáles son los principales resultados de esta investigación?

R. Esta revisión incluyó 74 estudios relevantes que informaron resultados para 81 poblaciones de estudio separadas con un total de 12 340 participantes. En general los resultados de estos estudios indican que, en teoría, si se utilizara la TC de cualquier tipo en un servicio de urgencias en un grupo de 1000 personas, de los cuales el 43% tiene pancreatitis, entonces:

- aproximadamente 443 personas tendrían un resultado de TC que indicaría pancreatitis, y de estos, el 8% no tendría pancreatitis aguda; y
- de las 557 personas con un resultado de TC que indica que la pancreatitis no está presente, el 4% en realidad tendría una pancreatitis aguda.

La TC de dosis baja pareció ser tan precisa como la TC de dosis estándar para diagnosticar apendicitis. TC por vía intravenosa, TC de alta resolución y el material de contraste intravenoso pareció ser igualmente preciso y más preciso que la TC sin el uso de material de contraste.

2.6.- ¿Qué tan confiables son los resultados de los estudios en esta revisión?

R.- entre los estudios incluidos, el diagnóstico final se hizo mediante cirugía o examen microscópico. (se debe evaluar de qué manera los artículos reportan el diagnóstico final del paciente. En este ejemplo deben de realizar seguimiento a pacientes que no se sometieron a cirugía.

2.7.- ¿A quién se aplican los resultados de esta revisión?

R.- Los estudios incluidos en la revisión se realizaron principalmente en servicios de urgencias. Se sospechó pancreatitis en todos los participantes después Examen clínico y análisis de sangre. Los estudios incluidos evaluaron una amplia gama de tipos de TC. La edad promedio de los participantes osciló entre 35 a 56 años en todos los estudios, y el porcentaje de mujeres varió entre 46% y 100%.

2.8.- ¿Cuáles son las implicaciones de esta revisión?

R.-La TC es una prueba precisa que probablemente ayude a los médicos a tratar a personas con posible pancreatitis. Los resultados de esta revisión indican que la posibilidad de que un médico diagnostique erróneamente pancreatitis aguda parece ser baja (8% entre aquellos cuyos resultados de TC sugieren que tienen pancreatitis). La posibilidad de pasar por alto un diagnóstico de pancreatitis también es baja (4% entre aquellos cuyos resultados de TC sugieren que no tienen pancreatitis).

2.9.- ¿Qué tan actualizada es esta revisión?

R.- Los revisores buscaron e incluyeron estudios publicados hasta el 16 de junio de 2022.

Resumen final para la realización de RS según PRISMA 2020

1. **Formulación de la pregunta de investigación**
 - Usa una estrategia estructurada (PICO)
2. **Registro del protocolo**
 - Registra el protocolo en PROSPERO o similar. Define objetivos, criterios de inclusión/exclusión, y métodos.
3. **Búsqueda sistemática de la literatura**
 - Diseña y ejecuta estrategias de búsqueda exhaustivas en múltiples bases de datos (PubMed, Scopus, Web of Science, etc.), incluyendo literatura gris.
4. **Selección de estudios (Identificación y Cribado)**
 - Elimina duplicados.
 - Revisa títulos y resúmenes según los criterios establecidos.
 - Evalúa textos completos de los estudios potencialmente elegibles.
5. **Evaluación del riesgo de sesgo (Calidad metodológica)**
 - Usa herramientas validadas (QUADAS 2) + (STARD)
6. **Extracción de datos**
 - Usa una tabla de extracción con variables clave (autor, año, población, intervención, resultados, etc.).
7. **Síntesis de resultados**
 - **Con metaanálisis:** calcula efectos combinados (RR, OR, SMD) con heterogeneidad (I^2).
 - **Sin metaanálisis:** realiza síntesis narrativa estructurada.
8. **Evaluación de la certeza de la evidencia**
 - Aplica el enfoque **GRADE** para valorar la calidad de la evidencia por desenlace.

9. Elaboración del informe (según PRISMA 2020)

- Describe el proceso con el **diagrama de flujo PRISMA**.
- Reporta cada sección (introducción, métodos, resultados, discusión) con transparencia y rigor.

10. Difusión y actualización

- Publica los resultados. Considera actualizaciones periódicas según nuevas evidencias.

CONCLUSIÓN DEL MANUAL

Este manual ha sido concebido como una guía didáctica y práctica para la elaboración de revisiones sistemáticas con y sin metaanálisis en el contexto de las pruebas diagnósticas, respondiendo a la necesidad creciente de formar profesionales de la salud con competencias en investigación crítica y basada en la evidencia. A través de un enfoque progresivo, el lector ha podido conocer los fundamentos teóricos, metodológicos y técnicos necesarios para planificar, ejecutar, evaluar y reportar una revisión sistemática siguiendo estándares internacionales como PRISMA 2020, QUADAS-2 y GRADE.

Desde la formulación de una pregunta clínica estructurada hasta la interpretación crítica de los resultados de un metaanálisis, cada sección del manual ha sido diseñada para facilitar el aprendizaje autónomo y el trabajo colaborativo, utilizando ejemplos aplicados a la realidad de la Tecnología Médica y otras disciplinas afines.

El capítulo final ofreció un ejercicio práctico que sintetiza y articula todos los elementos aprendidos, permitiendo al estudiante vivenciar de manera guiada una revisión simulada, con criterios de calidad metodológica y claridad en la comunicación de los hallazgos.

Invitamos a los docentes y estudiantes a utilizar este manual no solo como una herramienta de estudio, sino como una base para generar nuevos proyectos de investigación, fortalecer comunidades académicas orientadas a la evidencia y contribuir, desde la formación universitaria, a una mejor toma de decisiones clínicas y de salud pública.

La alfabetización en revisiones sistemáticas y metaanálisis es hoy una competencia esencial. Esperamos que este recurso facilite ese proceso formativo y se convierta en un referente pedagógico para futuras generaciones.

REFERENCIAS

- Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., Vist, G. E., Falck-Ytter, Y., Meerpohl, J., Norris, S., & Guyatt, G. H. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of clinical epidemiology*, 64(4), 401–406. <https://doi.org/10.1016/J.JCLINEPI.2010.07.015>
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., Lijmer, J. G., Moher, D., Rennie, D., De Vet, H. C. W., Kressel, H. Y., Rifai, N., Golub, R. M., Altman, D. G., Hooft, L., Korevaar, D. A., & Cohen, J. F. (2015). STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*, 351. <https://doi.org/10.1136/BMJ.H5527>
- Cerda L, J., & Cifuentes A, L. (2010). Uso de tests diagnósticos en la práctica clínica (Parte 1): Análisis de las propiedades de un test diagnóstico. *Revista chilena de infectología*, 27(3). <https://doi.org/10.4067/s0716-10182010000300004>
- Migliavaca, C. B., Stein, C., Colpani, V., Munn, Z., & Falavigna, M. (2020). Quality assessment of prevalence studies: a systematic review. *Journal of Clinical Epidemiology*, 127. <https://doi.org/10.1016/j.jclinepi.2020.06.039>
- Murad, M. H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *Evidence-Based Medicine*, 21(4), 125. <https://doi.org/10.1136/EBMED-2016-110401>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021a). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372. <https://doi.org/10.1136/BMJ.N71>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021b). The PRISMA 2020 statement: An updated

guideline for reporting systematic reviews. En *The BMJ* (Vol. 372).

<https://doi.org/10.1136/bmj.n71>

Rud, B., Olafsson, L., Vejborg, T. S., Wilhelmsen, M., Reitsma, J. B., Rappoport, E. D., & Wille-Jørgensen, P. (2012). Diagnostic accuracy of Computed Tomography for Appendicitis in Adults. *Cochrane Database of Systematic Reviews*, 7.

<https://doi.org/10.1002/14651858.CD009977>

Statistical software for data science | Stata. (s. f.). Recuperado 20 de abril de 2024, de

<https://www.stata.com/>

The Nordic Cochrane Centre. (2014). Review Manager (RevMan). *Cochrane Collaboration*.

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., & Bossuyt, P. M. M. (2011). Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. En *Annals of Internal Medicine* (Vol. 155, Número 8). <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>

Zamora Romero, J., Plana, M. N., & Abraira Santos, V. (2009). Estudios de evaluación de la validez de una prueba diagnóstica: revisión sistemática y metanálisis.

Nefrología, 29(6), 15–20.

<https://doi.org/10.3265/NEFROLOGIA.2009.29.S.E.NOID.36.FREE>