Máster en Salud Pública 2018-2019

Trabajo Fin de Máster

Sesgos en ensayos clínicos aleatorizados altamente citados publicados en revistas médicas con alto factor de impacto

Alumno: Francisco Román Calderón

Tutor: Ferrán Catalá López

Madrid, 7 de julio de 2019

RESUMEN

Introducción y objetivo: Evaluar y caracterizar el riesgo de sesgos existentes en ensayos clínicos aleatorizados publicados en revistas de alto impacto que hayan sido altamente citados.

Métodos: Análisis transversal de los artículos publicados en las revistas médicas con un factor de impacto > 10 (según el Journal Citation Reports publicado en 2017). Se realizaron búsquedas en PubMed/MEDLINE (desde su inicio hasta el 31 de diciembre de 2017) para todos los artículos relacionados con ensayos clínicos aleatorizados y controlados de intervenciones sanitarias con al menos 500 citas. Se extrajo información general de las características de los estudios y se evaluó el riesgo de sesgo utilizando la herramienta Cochrane para evaluar ensayos clínicos (7 dominios evaluados como riesgo alto, riesgo bajo o poco claro). Se realizaron análisis descriptivos con conteos de frecuencias, proporciones y diferencia de proporciones para una muestra aleatoria de estudios.

Resultados: Se evaluaron 40 ensayos clínicos con al menos 500 citas. Los artículos fueron publicados principalmente en cinco revistas (*The New England Journal of Medicine*: n=27; 67,5%; y *The Lancet*: n=9; 22,5%). La mitad (n=19; 47,5%) no indicó la disponibilidad del protocolo del estudio. Tan solo un 5% (n=2) fue publicado en acceso abierto. Principalmente se evaluaron intervenciones farmacológicas (n=22; 55%). Hubo gran diversidad en las indicaciones o enfermedades estudiadas. El 95% (n=38) de los estudios excluyó a pacientes con comorbilidad/multimorbilidad. Se clasificó el riesgo general de sesgo de los estudios como: “Alto riesgo” (al menos un dominio tenía alto riesgo) en 31 estudios (77,5%), “Poco claro” (al menos un dominio tenía riesgo incierto y ningún dominio presentaba alto riesgo) en 7 estudios (17,5%), y “Bajo riesgo” (todos los dominios tenían un bajo riesgo de sesgo) en 2 estudios (5%).

Conclusión: Los ensayos clínicos altamente citados que se publican en revistas de alto impacto no están exentos de potenciales sesgos. En qué medida estos sesgos pueden afectar a la práctica clínica necesita ser más estudiado.

Palabras clave: riesgo de sesgo, ensayo clínico aleatorizado, factor de impacto, metainvestigación, revisión sistemática, Colaboración Cochrane.

PREFACIO Y AGRADECIMIENTOS

Hasta donde me alcanza la memoria, siempre me ha interesado el por qué de las cosas, y especialmente de aquellas basadas en un juicio previo. Qué hace que algo sea bueno o malo, correcto o incorrecto, válido o inapropiado.

Pienso que en un mundo cada vez más regido por las opiniones de los expertos (y los no tan expertos), el método científico es un mástil al que asirse mientras dure la tempestad. Sin embargo, este no se encuentra exento de problemas, y no son pocos los que escudándose en el mismo, esgrimen argumentos, dan o implantan recomendaciones, sin pararse a contemplar si estas se han alcanzado de una forma lícita o incluso a sabiendas de que no lo han hecho.

Por eso con este trabajo quería aportar mi granito de arena para ayudar a identificar parte de esos posibles problemas, y de esa forma estar cada vez más cerca de ese concepto tan abstracto y a la vez tan atractivo de “la verdad”.

No quería tampoco dejar pasar la posibilidad de mostrar mis más sinceros agradecimientos:

A mi tutor, por su constante atención, comprensión, consejo y ayuda, sin las cuales este trabajo no hubiera llegado a buen puerto.

A la Escuela Nacional de Sanidad, por brindarme la oportunidad de iniciar mi carrera profesional de la mejor forma posible.

A mis compañeros, por su compañía y amistad, sin las cuales este último año no hubiera resultado una experiencia igual de positiva.

A mi pareja, por estar siempre ahí y hacerme feliz a cada instante que paso con ella.

A mi familia, por su incansable apoyo, sin el cual no estaría hoy donde estoy.

ÍNDICE

1. Introducción……..4
2. Hipótesis y objetivos……..7
3. Métodos……8
4. Resultados…….12
5. Discusión……18
6. Conclusiones……..…22
7. Bibliografía……..……...…23
8. Apéndice……26

Lista de tablas

- Tabla 1. Herramienta Cochrane para evaluar el riesgo de sesgos……...6
- Tabla 2. Características de los estudios incluidos……14
- Tabla 3. Riesgos de sesgos de los estudios incluidos……17
- Tabla 1A. Estrategia de búsqueda en PubMed/MEDLINE……26
- Tabla 2A. Revistas biomédicas incluidas con factor de impacto superior a 10…….……..…27
- Tabla 3A. Consideraciones metodológicas en la evaluación del riesgo de sesgo……..……...28
- Tabla 4A. Listado de los estudios incluidos con sus correspondientes citas……34

Lista de figuras

- Figura 1. Selección de artículos. Diagrama de flujo……..12
- Figura 1A. Evaluación del riesgo de sesgo a nivel de estudio..……..36
- Figura 2A. Resumen de la evaluación global del riesgo de sesgo……..37

Lista de abreviaturas

Ensayo clínico aleatorizado	ECA
Risk of Bias	*RoB*

1. INTRODUCCIÓN

Cuando se diseñan, ejecutan y presentan de manera apropiada, los ensayos clínicos aleatorizados (ECA) pueden informar la práctica clínica. Si no lo hacen, se vuelven éticamente cuestionables. Uno de los mayores problemas que se pueden presentar son los sesgos [1-5] y la falta de validez externa o de generalización de sus resultados, un desajuste entre la población estudiada y los pacientes del mundo real [6-8]. Para que los ECA formen parte de la base de la evidencia, por ejemplo, para establecer recomendaciones en guías de práctica clínica, es conveniente que se diseminen y publiquen de forma adecuada. Los ECA publicados en las principales revistas médicas son muy citados y suelen tener un papel fundamental en la práctica clínica.

Algunos trabajos previos han evaluado la calidad y el riesgo de sesgos de publicaciones de ECA en diversos contextos [1,4,5]. Por ejemplo, un pequeño estudio [5] publicado recientemente analizó una muestra de 10 ECA altamente citados (6.500 o más citaciones) poniendo de manifiesto potenciales sesgos que no fueron reconocidos por los investigadores de los estudios primarios. Por otro lado, estudios previos han sugerido que los pacientes que presentan múltiples enfermedades (comorbilidad/multimorbilidad) estarían infrarrepresentados en los ECA publicados [8].

Sesgos de los estudios

Se define como **sesgo** al error sistemático no provocado como un hecho aleatorio o por efecto del azar, que conlleva una desviación del efecto verdadero de una intervención, haciendo que este se sobrestime o infraestime. Los sesgos pueden producirse en cualquier etapa de un ECA, desde su planificación a la notificación de resultados. El *Manual Cochrane de revisiones sistemáticas de intervenciones* [9,10] clasifica los sesgos presentes en ECA de la siguiente manera:

- Sesgo de selección: hace referencia a las diferencias sistemáticas entre las características de los participantes que constituyen los grupos que componen el estudio.

- Sesgo de realización: se refiere a la presencia de diferencias sistemáticas entre los grupos del estudio en relación a la atención que se les proporciona, o en la exposición a factores diferentes de la intervención de interés.
- Sesgo de detección: en relación a si existieron diferencias sistemáticas entre los grupos de cara a cómo se obtuvieron los resultados.
- Sesgo de desgaste: si se produjeron diferencias sistemáticas en los abandonos o exclusiones de los participantes entre los grupos de un estudio.
- Sesgo de notificación: hace referencia a las diferencias entre los resultados preespecificados del estudio y los presentados a la finalización del mismo.

En 2005 varios miembros de la Colaboración *Cochrane* comenzaron a desarrollar la herramienta Cochrane para evaluar el riesgo de sesgos (o *"Cochrane Risk of Bias [RoB] Tool")*, con el objetivo de hacer el proceso de evaluación de riesgos de sesgos (parte fundamental de una revisión sistemática) más claro y preciso. En una primera fase, los autores desarrollaron criterios detallados para cada elemento incluido en la herramienta, dando lugar a una primera versión de la misma, que fue publicada en 2008 y recomendada como método de la Colaboración *Cochrane*.

En una fase de evaluación posterior, llevada a cabo en 2009, se recogieron las opiniones de los autores que habían utilizado la herramienta, y se revisó aquella primera versión, dando lugar a la publicación de la herramienta definitiva en 2011 [10]. Se propone utilizar la herramienta Cochrane para Riesgo de Sesgos en el presente trabajo con el objetivo de evaluar el riesgo de sesgo en una muestra aleatoria de artículos de ECA altamente citados y publicados en revistas biomédicas con alto factor de impacto. Brevemente, la herramienta se compone de 7 dominios, cada uno de los cuales está dedicado a evaluar el riesgo de los diferentes sesgos que se pueden producir a lo largo de una investigación (ver Tabla 1).

Tabla 1. Herramienta Cochrane para evaluar el riesgo de sesgos.

Dominio	Descripción	Tipo de sesgo
Sesgo de selección		
Generación de la secuencia	Describir el método utilizado para generar la secuencia de asignación con detalle suficiente para permitir una evaluación de si la misma produjo grupos comparables.	Sesgo de selección (asignación sesgada a las intervenciones) a causa de una generación inadecuada de la secuencia de aleatorización.
Ocultamiento de la asignación	Describir el método utilizado para ocultar la secuencia de asignación con detalle suficiente para determinar si las asignaciones a la intervención se podían prever antes o durante el reclutamiento.	Sesgo de selección (asignación sesgada a las intervenciones) a causa de una ocultación inadecuada de las asignaciones antes de asignarlas
Sesgo de realización		
Cegamiento de los participantes y del personal	Describir todas las medidas utilizadas, si se utilizó alguna, para cegar a los participantes y al personal del estudio al conocimiento de qué intervención recibió un participante. Proporcionar cualquier información con respecto a si el cegamiento propuesto fue efectivo.	Sesgo de realización a causa del conocimiento por parte de los participantes y del personal durante el estudio de las intervenciones asignadas.
Sesgo de detección		
Cegamiento de los evaluadores del resultado	Describir todas las medidas utilizadas, si se utilizó alguna, para cegar a los evaluadores del resultado del estudio al conocimiento de qué intervención recibió un participante. Proporcionar cualquier información con respecto a si el cegamiento propuesto fue efectivo	Sesgo de detección a causa del conocimiento por parte de los evaluadores de los resultados de las intervenciones asignadas.
Sesgo de desgaste		
Datos de resultado incompletos	Describir la compleción de los datos de resultado para cada resultado principal, incluidos los abandonos y las exclusiones del análisis. Señalar si se describieron los abandonos y las exclusiones, los números en cada grupo de intervención (comparados con el total de participantes asignados al azar), los motivos de las deserciones/exclusiones cuando se detallaron, y cualquier reinclusión en los análisis realizada por los revisores.	Sesgo de desgaste a causa de la cantidad, la naturaleza o el manejo de los datos de resultado incompletos.
Sesgo de notificación		
Notificación selectiva de los resultados	Señalar cómo los revisores examinaron la posibilidad de la notificación selectiva de los resultados, y qué encontraron	Sesgo de notificación a causa de la notificación selectiva de los resultados.
Otros sesgos		
Otras fuentes de sesgo	Señalar alguna inquietud importante acerca del sesgo no abordada en los otros dominios del instrumento. Si en el protocolo de la revisión se preespecificaron preguntas/ítems particulares, se deberían proporcionar las respuestas para cada pregunta/ítem	Sesgo debido a otros problemas no abordados en los apartados anteriores.

2. HIPÓTESIS Y OBJETIVOS

Hipótesis

Los artículos originales de ECA altamente citados publicados en revistas médicas con alto factor de impacto, presentan ciertas deficiencias en alguna o varias de sus fases de diseño, realización, análisis y/o notificación de resultados que pueden representar potenciales sesgos.

Desde que se publicara en 2011 la herramienta Cochrane para evaluar el riesgo de sesgo, la magnitud y el tipo de sesgos podría haber variado a lo largo del tiempo.

Objetivo principal

- Evaluar y caracterizar los sesgos existentes en ECA publicados en revistas de alto impacto que hayan sido altamente citados.

Objetivos secundarios:

- Llevar a cabo un análisis descriptivo de una muestra aleatoria de ECA altamente citados.
- Comparar el tipo y la magnitud del riesgo de sesgos en dos cortes transversales, antes de 2011 (2005-2010) y después de 2011 (2012-2017), coincidiendo con la fecha de publicación de la herramienta Cochrane para evaluar el riesgo de sesgos [10].
- Analizar la validez externa o generalización, a través de la existencia de multimorbilidad (o comorbilidad) [7,8] como criterios de inclusión o exclusión de los ECA evaluados.

3. MÉTODOS

Se ha llevado a cabo un análisis transversal de la literatura utilizando métodos sistemáticos de metainvestigación (investigación de la investigación) [12,13].

Se desarrolló un breve protocolo antes del inicio de este estudio. El protocolo se puede adquirir contactando al autor del trabajo.

No se requirió aprobación ética para la realización de este trabajo.

Búsqueda de la literatura

En marzo de 2019, se llevó a cabo una búsqueda sistemática en la base de datos MEDLINE a través de PubMed, con el objetivo de identificar los artículos de ECA publicados hasta el 31 de diciembre de 2017 en revistas biomédicas con alto factor de impacto. Para ello se empleó un filtro metodológico para identificar ensayos clínicos validado por Haynes et al. (con un 97% de especificidad, 93% de sensibilidad) [11] y los nombres de las principales revistas biomédicas que publican ECA. En el apéndice (Tabla 1A) se presenta la estrategia de búsqueda completa para facilitar su reproducibilidad.

Criterios de inclusión/exclusión

Se incluyeron artículos originales de ECA que evaluaron intervenciones sanitarias en población adulta, habían sido publicados en revistas biomédicas con un factor de impacto mayor de 10, de acuerdo al *Journal Citation Reports* [JCR] 2017 (Apéndice: Tabla 2A) y habían recibido al menos 500 citaciones en el momento de la búsqueda, de acuerdo con la información recogida en la *Web of Science*. Asimismo, se excluyeron los ECA en población pediátrica, análisis de subgrupos, análisis secundarios (*post-hoc*), las revisiones narrativas, revisiones sistemáticas, metanálisis, cartas, comentarios y artículos periodísticos.

Selección de la muestra

No se utilizaron métodos estadísticos para predeterminar el tamaño de la muestra. Se realizó una selección aleatoria de 40 artículos de ECA distribuidos en dos periodos (1:1) considerando la fecha de publicación en 2011 de la herramienta *Cochrane para evaluar el riesgo de sesgos* (grupo 1: 20 artículos publicados entre el 1 de enero de 2005 y el 31 de diciembre de 2010; grupo 2: 20 artículos publicados entre el 1 de enero de 2012 y el 31 de diciembre de 2017).

Recogida de información

Antes de proceder a la recogida de información, se llevó a cabo una prueba piloto que consistió en la revisión de 5 artículos (que posteriormente se incluyeron en la evaluación final), con los objetivos de familiarizarse con el manejo de la herramienta *Cochrane para evaluar el riesgo de sesgo* e *i*dentificar y subsanar posibles discrepancias en la extracción de datos de acuerdo con los criterios preestablecidos en el protocolo de la revisión.

Para cada artículo, un revisor (FRC) llevó a cabo la lectura crítica de los artículos originales, y extrajo la siguiente información: autor principal del artículo, revista y año de publicación, publicación en acceso abierto del artículo, política editorial y derechos de autor, disponibilidad del protocolo de estudio (y ubicación del protocolo en caso de estar disponible), enfermedad o indicación principal estudiada (de acuerdo a la clasificación internacional de enfermedades, 10ª edición), objetivo principal del estudio, tipo de objetivo principal (resultados importantes para los pacientes [p.ej., mortalidad o calidad de vida] o resultados intermedios/subrogados [p.ej., medidas de laboratorio]), naturaleza de la fuente de financiación del estudio (p.ej., financiación pública, privada, mixta), si existieron potenciales conflictos de interés, diseño del estudio (p.ej., grupos paralelos, cruzado, tipo cluster), intervención llevada a cabo en el estudio (p.ej., fármaco, procedimiento quirúrgico, terapia física), comparador empleado en el estudio, número de centros en los que se realizó el ECA, tamaño muestral del estudio, tiempo de seguimiento, la existencia de criterios de exclusión debidos a la edad, la existencia de criterios de exclusión debidos a la presencia de comorbilidades, entre otras.

De cara a contar con la mayor cantidad de información disponible, en cada caso se llevó a cabo una búsqueda exhaustiva del artículo publicado a texto completo, de los apéndices suplementarios del artículo, del protocolo del estudio, así como de los registros de ensayos clínicos en las bases de datos (como por ejemplo: Clinicaltrials.gov, International Clinical Trials Registry Platform, EU Clinical Trials Register y ISRCTN registry).

Tras la extracción de las variables descriptivas se procedió a la evaluación de los riesgos de sesgos en cada uno de los artículos. Un revisor (FRC) evaluó el riesgo de sesgo de cada estudio incluido. En caso de duda, se discutió cualquier desacuerdo entre dos revisores, y se emitió un juicio final si no se llegaba a un consenso. Los estudios se evaluaron utilizando la herramienta Cochrane para evaluar el riesgo de sesgo en ensayos clínicos. La herramienta incluye los siguientes dominios: generación de la secuencia, ocultamiento de la asignación, cegamiento de los participantes y el personal, cegamiento de los evaluadores de resultados, datos de resultados incompletos, notificación selectiva de resultados y otras fuentes de sesgo. Se evaluó cada dominio como “Bajo riesgo”, “Alto riesgo” o “Riesgo poco claro” de sesgo. Se clasificó el riesgo general de sesgo como “Bajo riesgo” si todos los dominios habían obtenido un bajo riesgo de sesgo, como “Alto riesgo” si al menos un dominio tenía un alto riesgo de sesgo, o “Riesgo poco claro” si al menos un dominio tenía un riesgo incierto de sesgo y ningún dominio presentaba alto riesgo. Esta regla está especificada por la herramienta Cochrane para evaluar el riesgo de sesgo [9,10]. Para más detalles, ver Tabla 3A en el apéndice.

La información fue recogida mediante hojas de datos diseñadas en Microsoft Excel (Seattle, Microsoft Corportation, EE.UU.) y en RevMan 5.3 (Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, Dinamarca).

Análisis de datos

Se llevó a cabo un análisis descriptivo mediante recuento de frecuencias y porcentajes para variables categóricas. Los resultados fueron estratificados en función de su fecha de publicación (2005-2010; 2012-2017). Se realizaron comparaciones entre las variables basadas en las diferencias de proporciones entre

el primer y el segundo periodo, calculando además sus intervalos de confianza al 95% (IC95%).

Los análisis se realizaron con el programa estadístico Stata 15 (StataCorp LP, College Station, Texas, EE.UU) .

4. RESULTADOS

Búsqueda de artículos y selección de la muestra

Se identificaron 39.329 registros en PubMed/MEDLINE, de los cuales 858 artículos resultaron elegibles (651 para el periodo 2005-2010 y 2017 para el periodo 2012-2017). Se seleccionó una muestra aleatoria de 40 estudios (ver diagrama de flujo en Figura 1, y Tabla 4A en apéndice con el listado completo).

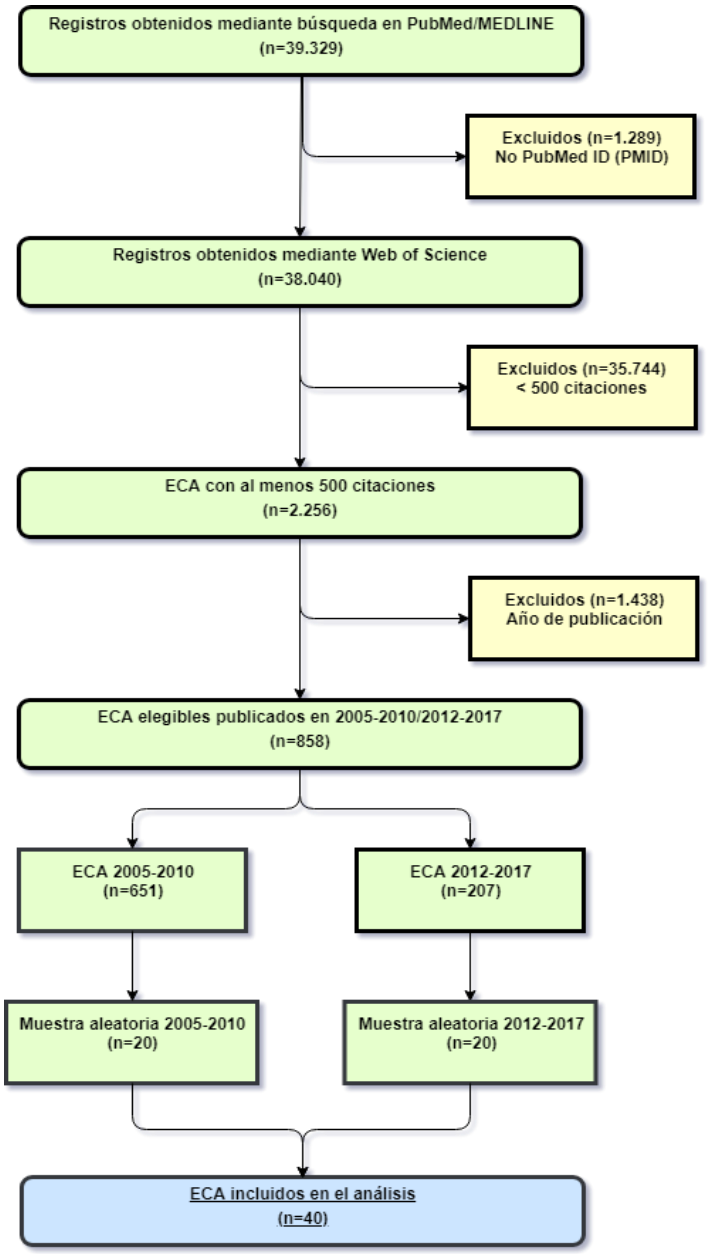


Figura 1. Selección de artículos. Diagrama de flujo

Características descriptivas

En la Tabla 2 se presentan las principales características de los 40 estudios incluidos. Los artículos altamente citados fueron publicados principalmente en cinco revistas con alto factor de impacto. Las principales revistas por número de artículos fueron: *The New England Journal of Medicine* (n=27; 67,5%) y *The Lancet* (n=9; 22,5%). Poco menos de la mitad (n=19; 47,5%) de los artículos no indicó la disponibilidad del protocolo del estudio. Tan solo un 5% (n=2) de los estudios fue publicado en acceso abierto, sin restricciones.

Prácticamente la totalidad de estudios empleó un diseño de grupos paralelos (n=39; 97,5%). Los estudios evaluaron intervenciones farmacológicas (n=22; 55%), no farmacológicas (n=14; 35%) o ambas (n=4; 10%). Hubo gran diversidad en las indicaciones o enfermedades estudiadas, con las neoplasias (n=17; 42,5%), las enfermedades del sistema circulatorio (n=7; 17,5%), las enfermedades endocrinas, nutricionales y metabólicas (n=5; 12,5%), y las enfermedades infecciosas y parasitarias (n=4; 10%) como las más comúnmente estudiadas.

La mitad de los estudios (n=20; 50%) consideró objetivos/variables de resultados importantes para los pacientes (como la mortalidad o la mejora de la calidad de vida). Sin embargo, en relación a los criterios de exclusión de los pacientes, un 95% (n=38) de los estudios excluyó a pacientes con comorbilidad/multimorbilidad.

La mayoría de los estudios contó con financiación privada: n=24; 60%; y los autores presentaban conflictos de interés en relación a la publicación del trabajo: n=28; 70% (Tabla 2).

Tabla 2. Características de los estudios incluidos

	2005-2010	(%)	2012-2017	(%)	Total	(%)
n	20		20		40	
Revista						
Arch Gen Psychiatry	1	5	0	0	1	2,5
JAMA	1	5	0	0	1	2,5
J Clin Oncol	1	5	0	0	1	2,5
Lancet	5	25	4	20	9	22,5
Lancet Oncol.	1	5	0	0	1	2,5
N Engl J Med	11	55	16	80	27	67,5
Acceso abierto						
Sí	1	5	1	5	2	5
No	19	95	19	95	38	95
Política editorial y derechos de autor						
Licencias exclusivas	19	95	19	95	38	95
Licencias no exclusivas	1	5	1	5	2	5
Protocolo						
Disponible	3	15	16	80	19	47,5
No disponible	17	85	4	20	21	52,5
Ubicación del protocolo						
Apéndice del artículo	2	66,7	16	100	18	94,7
Registro de EC	1	33,3	0	0	1	5,3
Enfermedad o indicación principal						
Ciertas enfermedades infecciosas y parasitarias	4	20	0	0	4	10
Embarazo, parto y puerperio	1	5	0	0	1	2,5
Enfermedades de la piel y el tejido subcutáneo	0	0	1	5	1	2,5
Enfermedades del sistema circulatorio	3	15	4	20	7	17,5
Enfermedades del sistema nervioso	1	5	1	5	2	5
Enfermedades del sistema respiratorio	0	0	1	5	1	2,5
Enfermedades endocrinas, nutricionales y metabólicas	1	5	4	20	5	12,5
Neoplasias	8	40	9	45	17	42,5
Trastornos mentales y del comportamiento	2	10	0	0	2	5
Tipo de objetivo principal						
Objetivo importante para el paciente	9	45	11	55	20	50
Objetivo subrogado	8	40	7	35	15	37,5
Ambos	3	15	2	10	5	12,5
Fuente de financiación						
Privada	10	50	14	70	24	60
Pública	5	25	6	30	11	27,5
Ambas	5	25	0	0	5	12,5

Tabla 2 (continuación). Características de los estudios incluidos

	2005-2010	(%)	2012-2017	(%)	Total	(%)
n	20		20		40	
Conflictos de interés						
Sí	12	60	16	80	28	70
No	8	40	4	20	12	30
Diseño del estudio						
Cruzado	1	5	0	0	1	2,5
Factorial	0	0	0	0	0	0
No controlado	0	0	0	0	0	0
Paralelo	19	95	20	100	39	97,5
Intervención						
Cuidados paliativos	1	3,7	0	0	1	2,1
Educación relacionada con la salud	1	3,7	0	0	1	2,1
Estilo de vida	1	3,7	1	4,8	2	4,2
Estrategia de seguimiento	0	0	1	4,8	1	2,1
Fármaco	11	40,7	15	71,4	26	54,2
Procedimiento diagnóstico	1	3,7	0	0	1	2,1
Procedimiento quirúrgico	5	18,5	2	9,5	7	14,6
Terapia física	5	18,5	1	4,8	6	12,5
Tiempo de tratamiento	1	3,7	1	4,8	2	4,2
Vacuna	1	3,7	0	0	1	2,1
Tipo de intervención						
Farmacológica	7	35	15	75	22	55
No farmacológica	9	45	5	25	14	35
Ambas	4	20	0	0	4	10
Comparador						
Comparador activo	3	15	11	55	14	35
Placebo/No tratamiento	17	85	9	45	26	65
Criterio de exclusión relacionado con la edad						
Sí	16	80	3	15	19	47,5
No	4	20	17	85	21	52,5
Edad del criterio de exclusión (años)						
< 18	13	56,5	11	52,4	24	54,5
18-65	2	8,7	6	28,6	8	18,2
> 65	8	34,8	4	19	12	27,3
Criterio de exclusión relacionado con la comorbilidad						
Sí	18	90	20	100	38	95
No	2	10	0	0	2	5
Tipo de Criterio de exclusión relacionado con la comorbilidad						
Explícito	4	22,2	0	0	4	10,5
Implícito	14	77,8	20	100	34	89,5

Riesgo de sesgos de los estudios

La Tabla 3, y las Figuras 1A y 2A (en el apéndice) proporcionan información sobre los resultados obtenidos tras las evaluaciones de los riesgos de sesgo de los estudios incluidos.

En la evaluación del riesgo de sesgo de selección mediante el dominio “Generación de la secuencia aleatoria” la proporción de artículos evaluados como “Bajo riesgo” fue superior (26/40;65%) a la de “Riesgo poco claro” (14/40; 35%). En este dominio no se alcanzó ninguna evaluación de “Alto riesgo”. En la comparativa entre periodos, la proporción de estudios evaluados como “Bajo riesgo” fue similar: primer periodo (13/20; 65%) frente al segundo periodo (13/20; 65%). En la evaluación del dominio “Ocultamiento de la asignación”, la mayoría de los estudios obtuvieron una valoración de “Bajo riesgo” (26/40; 65%) seguido de “Riesgo poco claro” (10/40; 25%). En este caso, se observaron diferencias entre periodos (alrededor del 30%) (ver Tabla 3).

En la evaluación del riesgo de sesgo de realización, valorado a través del dominio “Cegamiento de los participantes y del personal”, la mayoría de los estudios obtuvieron una calificación de “Alto riesgo” (27/40; 67,5%) seguidos de “Bajo riesgo” (10/40; 25%); mientras que solo 3/40 artículos fueron evaluados como “Riesgo poco claro”. En este dominio no existieron diferencias significativas entre periodos. Respecto al “Cegamiento de los evaluadores del resultado”, la mayoría de los estudios presentaron “Bajo riesgo” para el sesgo de detección (26/40; 65%). Un tercio de los estudios obtuvo una valoración de “Alto riesgo” para este mismo sesgo (13/40; 32,5%). En este caso, tampoco se observaron diferencias significativas entre periodos.

En la evaluación del riesgo de sesgo de desgaste, a través del dominio “Datos de resultado incompletos”, la mitad de los estudios (20/40;50%) fueron calificados como “Bajo riesgo”. Por otro lado, las evaluaciones de “Alto riesgo” (11/40; 27,5%) y “Riesgo poco claro” (9/40; 22,5%) presentaron proporciones similares. Comparando ambos periodos, no se observaron diferencias significativas entre periodos salvo para los dominios calificados con riesgo “Poco claro” con una diferencia del 25% (IC95%: 0,3- 50%) a favor del primer periodo.

El sesgo de notificación evaluado en el dominio “Notificación selectiva de los resultados” fue evaluado en la mayoría como “Bajo riesgo” (17/40; 42,5%) o “Riesgo poco claro” (16/40;40%). La valoración de “Alto riesgo” se presentó únicamente en 7/40 artículos (17,5%). Cuando enfrentamos los dos periodos, apreciamos que existen diferencias significativas en la proporción de los artículos valorados como “Bajo riesgo” (55%; IC95%: 30-80%), así como en los de “Riesgo poco claro” (60%; IC95%: 36-84%); a favor del segundo y del primer periodo, respectivamente.

Por último, se clasificó el riesgo general de sesgo de los estudios como: “Alto riesgo” (al menos un dominio tenía alto riesgo) en 31 estudios (77,5%), “Poco claro” (al menos un dominio tenía riesgo incierto y ningún dominio presentaba alto riesgo) en 7 estudios (17,5%), y “Bajo riesgo” (todos los dominios tenían un bajo riesgo de sesgo) en 2 estudios (5%).

Tabla 3. Riesgos de sesgos de los estudios incluidos

	Total (P1+P2)	*Total (%)*	2005-2010 (P1)	*2005-2010 (%)*	2012-2017 (P2)	*2012-2017 (%)*	Diferencia (P1 - P2)	*(I.C. 95%)*	
Generación de la secuencia									
Bajo riesgo	26	*65*	13	*65*	13	*65*	0%	*-30%*	*30%*
Alto riesgo	0	*0*	0	*0*	0	*0*	0%	*0%*	*0%*
Riesgo poco claro	14	*35*	7	*35*	7	*35*	0%	*-30%*	*30%*
Ocultamiento de la asignación									
Bajo riesgo	26	*65*	10	*50*	16	*80*	-30%	*-58%*	*-2%*
Alto riesgo	4	*10*	2	*10*	2	*10*	0%	*-19%*	*19%*
Riesgo poco claro	10	*25*	8	*40*	2	*10*	30%	*5%*	*55%*
Cegamiento de los participantes y del personal									
Bajo riesgo	10	*25*	3	*15*	7	*35*	-20%	*-46%*	*6%*
Alto riesgo	27	*67,5*	16	*80*	11	*55*	25%	*-3%*	*53%*
Riesgo poco claro	3	*7,5*	1	*5*	2	*10*	-5%	*-21%*	*11%*
Cegamiento de los evaluadores del resultado									
Bajo riesgo	26	*65*	11	*55*	15	*75*	-20%	*-49%*	*9%*
Alto riesgo	13	*32,5*	8	*40*	5	*25*	15%	*-14%*	*44%*
Riesgo poco claro	1	*2,5*	1	*5*	0	*0*	5%	*-5%*	*15%*
Datos de resultado incompletos									
Bajo riesgo	20	*50*	9	*45*	11	*55*	-10%	*-41%*	*21%*
Alto riesgo	11	*27,5*	4	*20*	7	*35*	-15%	*-42%*	*12%*
Riesgo poco claro	9	*22,5*	7	*35*	2	*10*	25%	*0,3%*	*50%*
Notificación selectiva de los resultados									
Bajo riesgo	17	*42,5*	3	*15*	14	*70*	-55%	*-80%*	*-30%*
Alto riesgo	7	*17,5*	3	*15*	4	*20*	-5%	*-28%*	*18%*
Riesgo poco claro	16	*40*	14	*70*	2	*10*	60%	*36%*	*84%*

5. DISCUSIÓN

El presente trabajo ha permitido evaluar y caracterizar los potenciales existentes en ensayos clínicos altamente citados en revistas médicas con alto factor de impacto. Generalmente, se encontró una proporción elevada de artículos publicados en revistas médicas británicas y estadounidenses (*The New England Journal of Medicine* y *The Lancet* representaron el 90% de artículos evaluados). Una hipótesis es que los diferentes patrones de publicación entre revistas pueden reflejar políticas editoriales y/o preferencias, con algunas revistas de medicina general más interesadas en la publicación de ensayos clínicos. En contraste, se observa que la gran mayoría de estos artículos se publica con restricciones de publicación y uso (muy pocos artículos altamente citados son de acceso abierto/Open Access) y, por lo tanto, los resultados de la investigación pueden no estar disponibles para una gran parte de la comunidad científica y la sociedad en general, incluidos clínicos (y pacientes) que deseen utilizar este conocimiento para informar su práctica clínica. Por otro lado, más de la mitad de los estudios evaluados no facilita información sobre el protocolo del estudio, a pesar de que la declaración CONSORT 2010 [14] recomienda indicar (ítem 24) “dónde puede accederse al protocolo completo del ensayo, si está disponible”.

Los resultados de este trabajo han permitido identificar que la mayoría de estudios analizados presenta alto riesgo de sesgos en al menos un dominio de la herramienta Cochrane para evaluar ensayos clínicos. Mientras que los dominios relacionados con el sesgo de selección (generación de la secuencia de aleatorización y ocultamiento de la asignación) sugieren resultados óptimos, al menos tres dominios brindan oportunidades: sesgo de realización, desgaste, y de notificación. Estos fueron los que más a menudo no se cumplieron y fueron consistentemente los predictores de calificaciones de calidad general más bajas (mayor riesgo de sesgo). Por ejemplo, a nivel riesgo de sesgo y dominios concretos, parece observarse un “Bajo riesgo/riesgo poco claro para dominios relacionados con sesgo de selección (generación de la secuencia/ocultamiento de la asignación) con valores inferiores al 5% para “Alto riesgo”, si bien para otros dominios es preocupante observase potenciales sesgos (2/3 estudios presentarían problemas en cegamiento de participantes y del personal), y en menor medida, sesgos de desgaste (27,5% estudios “Alto riesgo”) y de notificación selectiva (17,5% estudios

“Alto riesgo”). En este sentido, recientemente se han publicado los resultados del Proyecto COMPare [15,16] que evaluó el potencial sesgo de notificación selectiva de los resultados. Los autores examinaron los ECA publicados a lo largo de 6 semanas en 5 revistas de alto impacto (*Annals of Internal Medicine*, *The BMJ*, *JAMA*, *The Lancet*, and *The New England Journal of Medicine*). Solo nueve de 67 ensayos publicados informaron los resultados correctamente. Una cuarta parte no informó correctamente el resultado primario que se propusieron medir y el 45% no informó correctamente todos los resultados secundarios; otros añadieron nuevos resultados sin justificarlo. Además, cabe destacar que se observaron ciertas diferencias significativas en las evaluaciones de determinados dominios (p.ej., “Bajo riesgo” en el ocultamiento de la asignación) entre un periodo y otro; lo que anima a reflexionar sobre posibles tendencias temporales y hasta qué punto estas pueden estar motivadas por el uso de listas de verificación como SPIRIT 2013 [17] o CONSORT 2010 [18].

Por otro lado, los resultados de este estudio han permitido identificar que pocos ensayos clínicos altamente citados publicados en 2005-2017 incluyeron pacientes con múltiples enfermedades crónicas (comorbilidad/multimorbilidad). Estos hallazgos irían en la línea de trabajos previos. Por ejemplo, el estudio de Jadad et al [8] analizó 284 artículos publicados en revistas de alto impacto durante el periodo 1995-2010. Doscientos artículos (70%) mencionaron comorbilidad/multimorbilidad, y de ellos en 190 estudios (95%) su presencia afectó a la elegibilidad de los participantes. Por lo tanto, parece que siguen existiendo importantes lagunas de conocimiento en cuanto a comorbilidad y multimorbilidad en los ECA. Las limitaciones de las guías de práctica clínica y el tratamiento de enfermedades individuales están bien reconocidos en la literatura biomédica [7,8]. Los ensayos clínicos generalmente se realizan en poblaciones más o menos homogéneas, lo que puede impedir saber si los efectos de los tratamientos en personas con comorbilidad/multimorbilidad son equivalentes a los de pacientes con enfermedades individuales. La base de evidencia en las intervenciones para mejorar los resultados en las personas con multimorbilidad, por lo tanto, sigue siendo limitada. La consideración de las personas con multimorbilidad es esencial en el futuro diseño de estudios.

Algunas limitaciones pueden señalarse en este trabajo. La selección de estudios y la recogida de datos fue realizada por un solo revisor, por lo que es posible que existan potenciales errores en algunos datos. Del mismo modo, un único revisor evaluó los dominios de riesgos de sesgos definidos en la herramienta Cochrane. Para mitigar esta posible limitación, un segundo revisor realizó comprobaciones sobre un reducido número de trabajos. En caso de datos faltantes o poco claros, no se contactó con los autores de las publicaciones; aunque por otra parte, existe cierta evidencia que invita a tomar la información obtenida de este modo con cautela [19,20]. Esto puede llevar a que estudios bien realizados/ejecutados hayan sido calificados en algunos dominios como riesgo “Poco claro” ante la poca claridad en la descripción o ausencia de reporte. También cabe mencionar el reducido tamaño de la muestra de estudios evaluados limitando la potencia de los análisis entre los periodos analizados. Por último, se utilizó un único instrumento para evaluar el riesgo de sesgos de los estudios. Se escogió la herramienta Cochrane de riesgo de sesgos por ser la más utilizada en la actualidad, en lugar de otras que evalúan la calidad de los estudios a través de puntuaciones resumen (p.ej., escalada de Jadad) [21] ya que el uso de estas suele estar desaconsejado en revisiones sistemáticas y metanálisis al no estar respaldado por la evidencia empírica [9]. Sin embargo, a pesar de que la herramienta Cochrane proporciona una guía detallada sobre cómo usar la herramienta y recomienda el consenso entre dos revisores independientes, cabe mencionar que existe un juicio personal del/los evaluador/es, lo que puede generar fuentes de variabilidad. En este sentido, varios estudios han evaluado la reproducibilidad de la herramienta de riesgo de sesgo, que en ocasiones se ha mostrado deficiente [22-24].

Aunque la validez de los hallazgos está limitada por algunos de los aspectos anteriormente mencionados, los resultados invitan a reflexionar sobre los potenciales riesgos de sesgos y problemas en la generalización inadecuada de los resultados de los estudios realizados en las poblaciones estudiadas. Como futuras líneas de avance, se espera ampliar el trabajo presentado reforzando algunos aspectos. Principalmente, ampliando la muestra de estudios para reducir la incertidumbre y aumentar la precisión de las comparaciones en dos cortes transversales (p.ej., sobre una muestra aleatoria de 300 estudio) [25,26], realizando evaluaciones del riesgo de sesgo por duplicado y de forma independiente por al

menos 2 revisores, y evaluando sesgos para otros resultados (como sesgo de reporte selectivo de resultados para objetivos secundarios de seguridad de las intervenciones evaluadas), y contactando a los autores de los estudios en busca de información adicional (p.ej., acceso al protocolo del estudio o facilitar información para dominios con riesgo “Poco claro”).

6. CONCLUSIONES

Los ensayos clínicos altamente citados que se publican en revistas de alto impacto no están exentos de potenciales sesgos. Considerar los diversos dominios de la herramienta Cochrane para evaluar el riesgo de sesgos puede ayudar a identificar qué aspectos pueden fortalecerse, y permiten que la comunidad científica se centre en formas constructivas para mejorar el rigor y la integridad en la investigación que se publica en las revistas médicas más influyentes. Por otro lado, en qué medida estos sesgos pueden afectar a la práctica clínica necesita ser más estudiado.

7. BIBLIOGRAFÍA

1. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ; GRADE Working Group. What is "quality of evidence" and why is it important to clinicians? BMJ. 2008 May 3;336(7651):995-8.

2. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. BMJ. 2010 Mar 23;340:c723.

3. Hodkinson A, Kirkham JJ, Tudur-Smith C, Gamble C. Reporting of harms data in RCTs: a systematic review of empirical assessments against the CONSORT harms extension. BMJ Open. 2013 Sep 27;3(9):e003436.

4. Haidich AB, Birtsou C, Dardavessis T, Tirodimos I, Arvanitidou M. The quality of safety reporting in trials is still suboptimal: survey of major general medical journals. J Clin Epidemiol. 2011 Feb;64(2):124-35.

5. Dechartres A, Trinquart L, Atal I, Moher D, Dickersin K, Boutron I, Perrodeau E, Altman DG, Ravaud P. Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. BMJ. 2017 Jun 8;357:j2490.

6. Krauss A. Why all randomised controlled trials produce biased results. Ann Med. 2018 Jun;50(4):312-322

7. Muth C, Glasziou PP. Guideline recommended treatments in complex patients with multimorbidity. BMJ. 2015 Oct 2;351:h5145.

8. Jadad AR, To MJ, Emara M, Jones J. Consideration of multiple chronic diseases in randomized controlled trials. JAMA. 2011 Dec 28;306(24):2670-2.

9. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Disponible: www.cochrane-handbook.org.

10. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA; Cochrane Bias Methods Group;

Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011 Oct 18;343:d5928.

11. Haynes RB, McKibbon KA, Wilczynski NL, Walter SD, Werre SR; Hedges Team. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. BMJ. 2005 May 21;330(7501):1179.

12. Ioannidis JPA. Meta-research: Why research on research matters. PLoS Biol. 2018 Mar 13;16(3):e2005468.

13. Ioannidis JP, Fanelli D, Dunne DD, Goodman SN. Meta-research: Evaluation and Improvement of Research Methods and Practices. PLoS Biol. 2015 Oct 2;13(10):e1002264.

14. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340:c869

15. Goldacre B, Drysdale H, Dale A, Milosevic I, Slade E, Hartley P, Marston C, Powell-Smith A, Heneghan C, Mahtani KR. COMPare: a prospective cohort study correcting and monitoring 58 misreported trials in real time. Trials. 2019;20(1):118.

16. Goldacre B, Drysdale H, Marston C, Mahtani KR, Dale A, Milosevic I, Slade E, Hartley P, Heneghan C. COMPare: Qualitative analysis of researchers' responses to critical correspondence on a cohort of 58 misreported trials. Trials. 2019;20(1):124.

17. Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials. Ann Intern Med. 2013 Feb 5;158(3):200.

18. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010 Mar 23;340:c332.

19. Haahr MT, Hróbjartsson A. Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors. Clin Trials. 2006;3(4):360–5.

20. Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials. JAMA. 2004 May 26;291(20):2457.

21. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996 Feb;17(1):1-12.

22. Hartling L, Hamm MP, Milne A, Vandermeer B, Santaguida PL, Ansari M, Tsertsvadze A, Hempel S, Shekelle P, Dryden DM. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. J Clin Epidemiol. 2013 Sep;66(9):973-81.

23. Hartling L, Hamm M, Milne A, Vandermeer B, Santaguida PL, Ansari M, Tsertsvadze A, Hempel S, Shekelle P, Dryden DM. Validity and Inter-Rater Reliability Testing of Quality Assessment Instruments [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012 Mar. Available from http://www.ncbi.nlm.nih.gov/books/NBK92293/

24. Bertizzolo L, Bossuyt P, Atal I, Ravaud P, Dechartres A. Disagreements in risk of bias assessment for randomised controlled trials included in more than one Cochrane systematic reviews: a research on research study using cross-sectional design. BMJ Open. 2019 Apr 1;9(4):e028382.

25. Schriger DL, Altman DG, Vetter JA, Heafner T, Moher D. Forest plots in reports of systematic reviews: a cross-sectional study reviewing current practice. Int J Epidemiol. 2010 Apr;39(2):421-9.

26. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, Catalá-López F, Li L, Reid EK, Sarkis-Onofre R, Moher D. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. PLoS Med. 2016 May 24;13(5):e1002028.

8. APÉNDICE

Tabla 1A. Estrategia de búsqueda en PubMed/MEDLINE.

Búsqueda	Query
#1	*"N Engl J Med"[Journal] OR "Lancet"[Journal] OR "JAMA"[Journal] OR "J Am Med Assoc"[Journal] OR "BMJ"[Journal] OR "Br Med J"[Journal] OR "Br Med J (Clin Res Ed)"[Journal] OR "Ann Intern Med"[Journal] OR "JAMA Intern Med"[Journal] OR "Arch Intern Med"[Journal] OR "PLoS Med"[Journal] OR "Lancet Oncol"[Journal] OR "World Psychiatry"[Journal] OR "Lancet Neurol"[Journal] OR "J Clin Oncol"[Journal] OR "Eur Heart J"[Journal] OR "J Am Coll Cardiol"[Journal] OR "Lancet Infect Dis"[Journal] OR "Lancet Diabetes Endocrinol"[Journal] OR "Circulation"[Journal] OR "Lancet Respir Med"[Journal] OR "Gastroenterology"[Journal] OR "Gut"[Journal] OR "JAMA Oncol"[Journal] OR "Eur Urol"[Journal] OR "JAMA Psychiatry"[Journal] OR "Arch Gen Psychiatry"[Journal] OR "Am J Psychiatry"[Journal] OR "Circ Res"[Journal] OR "Hepatology"[Journal] OR "Am J Respir Crit Care Med"[Journal] OR "Blood"[Journal] OR "J Allergy Clin Immunol"[Journal] OR "Ann Rheum Dis"[Journal] OR "J Natl Cancer Inst"[Journal] OR "J Hepatol"[Journal] OR "Intensive Care Med"[Journal] OR "Diabetes Care"[Journal] OR "Ann Oncol"[Journal] OR "Leukemia"[Journal] OR "Lancet Psychiatry"[Journal] OR "Eur Respir J"[Journal] OR "Brain"[Journal] OR "JAMA Pediatr"[Journal] OR "Arch Pediatr Adolesc Med"[Journal] OR "JAMA Neurol"[Journal] OR "Arch Neurol"[Journal]*
#2	*(randomized controlled trial[Publication Type] OR (randomized[Title/Abstract] AND controlled[Title/Abstract] AND trial[Title/Abstract]))*
#3	*("0001/01/01"[PDAT] : "2017/12/31"[PDAT])*
#4	*#1 AND #2 AND #3*
#5	*(systematic review[title] OR review[publication type] OR meta-analy*[title] OR meta-analysis[publication type] OR pool analys*[title] OR pooled data [text word] OR letter [publication type] OR newspaper article [publication type])*
#6	***#4 NOT #5***

Tabla 2A. Revistas biomédicas incluidas con factor de impacto superior a 10.

Título de la revista	**Factor de impacto (JCR 2017)**
The New England Journal of Medicine	79,26
The Lancet	53,25
JAMA – the Journal of the American Medical Association	47,66
The BMJ – the British Medical Journal	23,56
JAMA Internal Medicine – antes, Archives of Internal Medicine	19,99
Annals of Internal Medicine	19,38
PLOS Medicine	11,67
Lancet Oncology	36,42
World Psychiatry	30
Lancet Neurology	27,14
Journal of Clinical Oncology	26,36
Lancet Infectious Diseases	25,15
European Heart Journal	23,42
JACC – *Journal of the American College of Cardiology*	16,83
Lancet Respiratory Medicine	21,47
JAMA Oncology	20,87
Gastroenterology	20,77
Lancet Diabetes & Endocrinology	19,31
Circulation	18,88
European Urology	17,58
Gut	17,02
JAMA Psychiatry –antes, Archives of General Psychiatry	16,64
American Journal of Respiratory and Critical Care Medicine	15,24
Lancet Psychiatry	15,23
Circulation Research	15,21
Blood	15,13
Journal of Hepatology	15,04
Intensive Care Medicine	15,01
Hepatology	14,08
Annals of Oncology	13,93
American Journal of Psychiatry	13,4
Diabetes Care	13,4
Journal of Allergy and Clinical Immunology	13,26
Annals of the Rheumatic Diseases	12,35
European Respiratory Journal	12,24
JAMA Neurology – antes, Archives of Neurology	11,46
JNCI – *Journal of the National Cancer Institute*	11,24
Brain	10,85
JAMA Pediatrics – antes, Archives of Pediatrics & Adolescent Medicine	10,77
Leukaemia	10,02

Tabla 3A. Consideraciones metodológicas en la evaluación del riesgo de sesgo.

Generación aleatoria de la secuencia

Para valorar el primer dominio se tuvo en cuenta si se ofrecía información de forma explícita sobre cual había sido el método empleado para obtener una secuencia de asignación. Aquellos basados en la asignación aleatoria simple obtuvieron la evaluación de "Bajo riesgo" (p.ej. generación de números aleatorios por ordenador, tabla de números aleatorios, lanzamiento de una moneda, barajar cartas o sobres, lanzar dados, sorteo de tarjetas).

También consideramos "Bajo riesgo" a aquellos procesos de asignación aleatoria limitada como la asignación en bloques, en la que los participantes de un estudio se van asignando a uno de los grupos de forma consecutiva hasta alcanzar un número predeterminado (p.ej. 4), momento en el que comienzan a ser asignados a otro grupo, al cual se asignarán el mismo número de participantes hasta completar el bloque y comenzar de nuevo la secuencia.

Por último, se evaluó como "Bajo riesgo" en el primer dominio a los estudios en los que se llevó a cabo una asignación aleatoria estratificada, siempre y cuando se ofreciera información suficiente sobre cuáles fueron los factores pronósticos empleados en la estratificación y se considerara que dichos factores habían sido suficientes para asegurar la creación de grupos comparables. Por otro lado, el empleo de métodos cuasi-aleatorios, como la asignación basada en la fecha de nacimiento de los participantes, la fecha de consulta o ingreso de los mismos, así como el número de historia clínica o cualquier otro tipo de identificador (no obtenido mediante generación aleatoria), métodos de alternancia o, por supuesto, asignación en función del criterio del médico, resultados de una prueba o preferencia del participante, fueron evaluados con "Alto riesgo". Todos los métodos no contemplados en los puntos anteriores, como la minimización, y más frecuentemente, la ausencia de información suficiente sobre el proceso de generación de la secuencia (p.ej. "se llevó a cabo un estudio aleatorizado", o "los participantes fueron asignados aleatoriamente"), se tradujo en una evaluación de "Riesgo poco claro".

Ocultación de la asignación

Se valoró si los participantes y/o las personas encargadas de reclutar a los mismos pudieron anticipar el grupo al que iban a ser asignados. Cabe destacar aquí que en este dominio se hace referencia a las técnicas empleadas para implementar la secuencia de asignación, no para generarla. De dicha forma, la generación de una secuencia aleatoria no siempre asegura un bajo riesgo de sesgo de detección, ya que si el participante o el investigador que asigna la intervención conocen *a priori* el grupo al que van a ser asignados, pueden modificar su decisión, dando lugar a diferencias sistemáticas entre los distintos grupos.

Es importante recalcar también que la ocultación de la asignación tiene lugar antes de que la intervención sea asignada (y por lo tanto aplicada), por lo que siempre se puede llevar a cabo si se emplean los métodos adecuados, incluso en aquellos casos en los que el cegamiento de los participantes y el personal no es posible.

Consideramos como adecuados aquellos sistemas que llevaron a cabo la asignación de forma centralizada, como los denominados *Interactive Voice Response System* (IVRS) o *Interactive Web Response System* (IWRS), o los más tradicionales métodos de asignación por teléfono o fax. También fueron considerados como "Bajo riesgo" el uso de sobres cerrados, oscuros y numerados de forma secuencial o envases del fármaco, numerados de forma secuencial y con la misma apariencia.

Por el contrario, los procedimientos de asignación en los que los participantes o los investigadores tuvieron la posibilidad de anticipar el grupo al que iban a ser asignados se evaluó como "Alto riesgo". Este sería el caso de esquemas de asignación al azar abiertos (p.ej. lista de números aleatorios colgada en la pared) o métodos descritos en el punto anterior que no contaron con la protección adecuada (p.ej. sobres no sellados o no opacos, o envases de fármaco distintos). Procedimientos basados en secuencias cuasi-aleatorias (p.ej. fecha de nacimiento o número de historia clínica) también fueron considerados como "Alto riesgo" de sesgo de detección.

Resulta especialmente importante en este punto la descripción minuciosa de los métodos, especialmente aquellos que no se llevaron a cabo de manera centralizada. Aquellos casos en los que no hubo información suficiente para evaluar al dominio como "Bajo riesgo" o "Alto riesgo" o en los que no se describieron con suficiente detalle los métodos empleados, obtuvieron una valoración de "Riesgo poco claro".

<u>Cegamiento de los participantes y del personal</u>

En el siguiente dominio se consideró el riesgo de sesgo de realización mediante la valoración del cegamiento de los participantes y el personal, entendiendo a estos últimos como los encargados de la asistencia sanitaria de los primeros, aunque esta se llevara a cabo en un ámbito distinto al correspondiente a la intervención del estudio.

Se ha demostrado que la falta de cegamiento en los ECA conlleva una sobrestimación del efecto de la intervención de hasta un 9%, por lo que se valoró con especial meticulosidad este dominio, atendiendo no solo a si hubo intención de cegar a los participantes y el personal, sino también a si este cegamiento fue posible. Resulta muy complicado, por ejemplo, cegar de forma efectiva intervenciones quirúrgicas; casos en los que los fármacos tenían apariencias o vías de administración, o pautas distintas; o casos en los que la naturaleza de las intervenciones difería completamente (p.ej. fármaco antidepresivo vs. terapia cognitiva).

Por lo tanto, de cara a la evaluación de este dominio se hace imprescindible la descripción detallada tanto de los métodos empleados para conseguir el cegamiento (términos como "doble ciego" o "ECA cegado" no son suficientes para valorar el riesgo de sesgo de realización, como de las propias intervenciones contempladas en el estudio. E incluso en el caso de que todo lo anterior fuese adecuado, llevamos a cabo una valoración final sobre la probabilidad de que el cegamiento acabara siendo efectivo.

Aquellos estudios denominados como “abiertos”, “desenmascarados”, “no cegados”, etc. fueron clasificados directamente como “Alto riesgo” de sesgo de realización.
Para simplificar esta tarea se desarrolló un árbol de toma de toma de decisiones que se encuentra incluido en el protocolo de esta revisión.

Cegamiento de los evaluadores de los resultados

A continuación se valoró el dominio referente al cegamiento de los evaluadores, entendiendo a estos como aquellas personas encargadas de medir uno o más resultados del estudio. Dado que el cegamiento de los evaluadores puede verse afectado por los mismos o incluso más factores que el de los participantes y el personal, la valoración de este dominio se llevó a cabo de forma tan minuciosa como el anterior.
De nuevo, la simple intención de cegar a los evaluadores no garantiza que esta se llevara a cabo o que fuera efectiva. Por ello se tuvo en cuenta, en primer lugar, quien o quienes fueron los encargados de evaluar los resultados. En el caso de que esta tarea recayera sobre los propios participantes (p.ej. resultados de calidad de vida) o el personal (p.ej. medición del tamaño de una placa de psoriasis), se atendió a la evaluación del dominio anterior. Si por el contrario, la evaluación corrió a cargo de personal no encargado de la atención sanitaria del participante o de una persona o comité independientes, se tuvo en cuenta si la evaluación requería un juicio por parte del evaluador; en caso negativo (p.ej. niveles de proteína c reactiva), se evaluó el dominio como “Bajo riesgo”; si por el contrario la evaluación sí requería un juicio por parte del evaluador (p.ej. valoración de una radiografía) se contempló si este juicio se realizó en base a unos criterios preespecificados (p.ej. escala RECIST para determinar la evolución de un tumor), o simplemente según el criterio del evaluador (p.ej. alta hospitalaria). En el primer caso se evaluó el dominio como “Bajo riesgo”, pero siempre contemplando la posibilidad de que el cegamiento no fuese efectivo. En el segundo caso se valoró si fue probable que el cegamiento se rompiera, y en función de esto se evaluó como “Alto riesgo” o “Bajo riesgo” de sesgo de detección.
Todos aquellos artículos en los que no se proporcionó suficiente información para valorar el riesgo de sesgo de detección como “Alto riesgo” o “Bajo riesgo” se evaluaron como “Riesgo poco claro”.
Al igual que en el dominio anterior, se adjunta en el protocolo un árbol de decisión empleado para valorar el cegamiento de los evaluadores.
Cabe destacar que existen ciertos resultados que pueden no verse afectados por el cegamiento de los evaluadores. En nuestro caso, consideramos que los resultados de “supervivencia” o “mortalidad” global o “por todas las causas” fueron independientes del cegamiento de los evaluadores y fueron evaluadas como “Bajo riesgo” en lo referente a este dominio.

Datos de resultado incompletos

En lo referente a la sección de resultados, prestamos atención al quinto dominio, relativo al sesgo de desgaste, en el que se valora si las pérdidas o exclusiones producidas a lo largo del estudio han podido influir sobre la estimación del efecto observado.

En este dominio se consideraron cuatro aspectos fundamentales: la magnitud de los datos faltantes, su distribución, los motivos que los propiciaron y la estrategia de análisis empleada.

En lo concerniente a la magnitud de los datos faltantes, tuvimos en cuenta si el resultado principal del estudio correspondía a una variable categórica o cuantitativa. En el primer caso se comparó la proporción de pérdidas y/o exclusiones de cada rama del estudio, con la probabilidad de ocurrencia del resultado principal en el grupo control. Para resultados cuantitativos se contempló en primer lugar si se habían realizado análisis de sensibilidad y si estos habían arrojado resultados similares a los obtenidos en el análisis principal. Si dichos análisis de sensibilidad no se habían realizado o no se aportaron en el artículo principal o su apéndice, se pasó a contemplar la proporción de pérdidas y/o exclusiones en cada rama del estudio respecto al número total de individuos aleatorizados en esa misma rama.

En cuanto a la distribución de los datos faltantes, se comparó la proporción de pérdidas y/o exclusiones del grupo/s de intervención, con la del grupo control.

A continuación se atendió a los motivos de los datos faltantes. En caso de que se ofrecieran, se valoró si era probable que estos estuvieran relacionados con el resultado principal y si eran similares o diferían entre los distintos grupos.

Por último, se tuvo en cuenta la estrategia de análisis empleado. Aquí podemos encontrarnos con cuatro aproximaciones distintas: análisis por intención de tratar (AIT), análisis de casos disponibles, análisis por protocolo, análisis por tratamiento recibido.

El AIT debía cumplir varios supuestos para que fuera considerado como tal, a saber: que el análisis incluyera a todos los participantes aleatorizados, que estos fueran incluidos en la rama del estudio a la que fueron asignados y que el resultado se midiera en todos los participantes aleatorizados. Dado que este último punto es virtualmente imposible para la mayoría de los ECA, se tuvo en cuenta los métodos de imputación o censura empleados de cara al análisis.

En el análisis de casos disponibles, se debía incluir a todos los participantes en los cuales se habían obtenido resultados y hacerlo en la rama del estudio a la que habían sido asignados.

Se consideró análisis por protocolo a aquel en el que únicamente se incluyó a los pacientes que habían cumplido con la intervención en el momento de finalización del estudio (o de realización del análisis).

Finalmente, se contempló como análisis por tratamiento recibido al que incluyó a los pacientes en el grupo de la intervención que habían recibido, independientemente del grupo al que habían sido asignados.

Como norma general, se evaluaron como “Bajo riesgo” a aquellos estudios en los cuales las pérdidas y/o exclusiones respondieron a las siguientes características:

- La proporción de pérdidas y/o exclusiones en cada una de las ramas del estudio, fue menor que la probabilidad de ocurrencia del resultado principal en el grupo control.
- La proporción de pérdidas y/o exclusiones en cada una de las ramas del estudio fue, como máximo, 1,5 veces la proporción de pérdidas y/o exclusiones del grupo control.
- Los motivos de las pérdidas y/o exclusiones no están relacionados con el resultado principal (p.ej. cambio lugar de residencia) y son similares en ambos grupos.
- Se llevó a cabo un AIT o un análisis de casos disponibles.

Por el contrario, y también como norma general, se evaluaron como “Alto riesgo” a aquellos estudios en los cuales las pérdidas y/o exclusiones respondieron a las siguientes características:

- La proporción de pérdidas y/o exclusiones en cada una de las ramas del estudio, fue mayor que la probabilidad de ocurrencia del resultado principal en el grupo control
- La proporción de pérdidas y/o exclusiones en cada una de las ramas del estudio fue, como mínimo, 2 veces la proporción de pérdidas y/o exclusiones del grupo control.
- Los motivos de las pérdidas y/o exclusiones están relacionados con el resultado principal (p.ej. efectos adversos) y son diferentes en ambos grupos.
- Se llevó a cabo un análisis por protocolo o un análisis por tratamiento recibido.

En aquellos casos en los que no se ofrecieron en detalle los motivos de las pérdidas y/o exclusiones, o información sobre los métodos de imputación o censura empleados en el AIT, así como toda la casuística no contemplada en los puntos anteriores, se valoró el riesgo de sesgo de desgaste como “Riesgo poco claro”.

En el protocolo de la revisión, disponible en el apéndice de este memoria, se detalla más específicamente cada uno de los puntos a tener en cuenta de cara a la valoración de este dominio, así como las combinaciones en las que se basaron las evaluaciones finales del mismo.

En cualquier caso, los criterios anteriormente mencionados serán tomados como una guía de evaluación de carácter orientativo, ya que debido a la distinta naturaleza de los resultados incluidos en esta revisión, se analizará cada estudio en su contexto particular.

<u>Notificación selectiva de los resultados</u>

Para valorar este dominio comparamos la información ofrecida en el artículo principal (y su apéndice en caso de estar disponible) referente al resultado principal con la preespecificada en el protocolo del estudio. En caso de que el protocolo del estudio no estuviera disponible, se recurrió a los registros de ensayos clínicos anteriormente mencionados; pero solo se tuvieron en cuenta aquellos que contaban información correspondiente a fechas anteriores al inicio de comienzo del estudio, considerando esta como la fecha en la que comenzó el reclutamiento de los participantes. Cuando ninguna de las anteriores fuentes de información estuvo disponible, evaluamos el riesgo de notificación selectiva como “Riesgo poco claro”.

Aquellos artículos en los que contamos con información previa válida, comparamos si los desenlaces preespecificados coincidían con los reportados en la sección de resultados, tanto en lo referente a los propios desenlaces como a las características de los mismos, p.ej. si se incluía la significación estadística, si coincidían los puntos temporales de los resultados, los instrumentos o escalas empleados para medir, los análisis estadísticos empleados, si solo se ofrecían resultados de subgrupos…etc. En general, cualquier desviación de los datos aportados de forma anterior al comienzo del estudio respecto a la ofrecida posteriormente en el artículo fue evaluada como "Alto riesgo" de sesgo de notificación. Si, por el contrario, ambas informaciones coincidieron, se evaluó el dominio como "Bajo riesgo".

Otros sesgos

A lo largo de la lectura del artículo, se valoró la posible existencia de otros sesgos no contemplados en los dominios anteriores.

Un punto a tener en cuenta en la evaluación de los dominios, es que todos aquellos que hacen referencia a los resultados ("Cegamiento de los participantes y del personal", "Cegamiento de los evaluadores de los resultados", "Datos de resultado incompletos" y "Notificación selectiva de los resultados") se valoraron únicamente para el resultado principal del estudio (*primary end point)*.

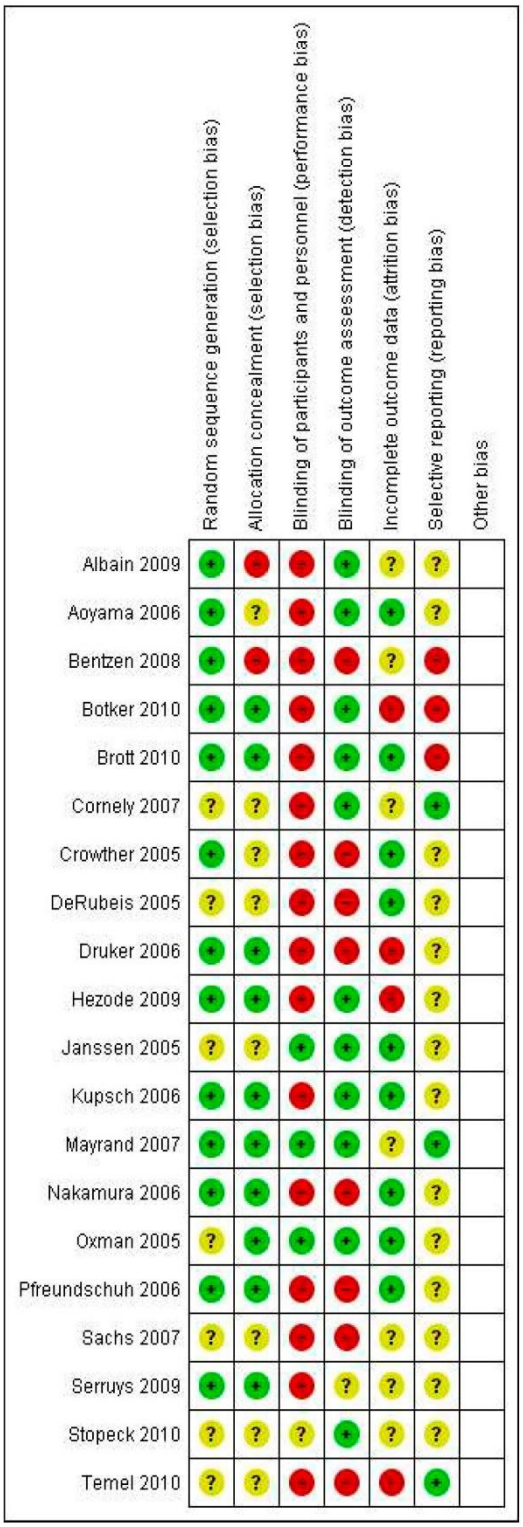
Tabla 4A. Listado de los estudios incluidos con sus correspondientes citas.

Núm.	Artículo	Total citas
1	Temel JS, Greer JA, Muzikansky A, Gallagher ER, Admane S, Jackson VA, et al. Early Palliative Care for Patients with Metastatic Non–Small-Cell Lung Cancer. N Engl J Med . 2010;363(8):733–42.	2855
2	Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. N Engl J Med. 2015;373(17):1627–39.	2612
3	Druker BJ, Guilhot F, O'Brien SG, Gathmann I, Kantarjian H, Gattermann N, et al. Five-Year Follow-up of Patients Receiving Imatinib for Chronic Myeloid Leukemia. N Engl J Med. 2006;355(23):2408–17.	2186
4	Serruys PW, Morice M-C, Kappetein AP, Colombo A, Holmes DR, Mack MJ, et al. Percutaneous Coronary Intervention versus Coronary-Artery Bypass Grafting for Severe Coronary Artery Disease. N Engl J Med. 2009;360(10):961–72.	2137
5	Flaherty KT, Robert C, Hersey P, Nathan P, Garbe C, Milhem M, et al. Improved Survival with MEK Inhibition in BRAF-Mutated Melanoma. N Engl J Med. 2012;367(2):107–14.	1767
6	Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS. Effect of Treatment of Gestational Diabetes Mellitus on Pregnancy Outcomes. N Engl J Med. 2005;352(24):2477–86.	1515
7	Marso SP, Daniels GH, Brown-Frandsen K, Kristensen P, Mann JFE, Nauck MA, et al. Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med. 2016;375(4):311–22.	1463
8	Brott TG, Hobson RW, Howard G, Roubin GS, Clark WM, Brooks W, et al. Stenting versus endarterectomy for treatment of carotid-artery stenosis. N Engl J Med. 2010;363(1):11–23.	1445
9	Baselga J, Campone M, Piccart M, Burris HA, Rugo HS, Sahmoud T, et al. Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. N Engl J Med. 2012;366(6):520–9.	1426
10	Oxman MN, Levin MJ, Johnson GR, Schmader KE, Straus SE, Gelb LD, et al. A Vaccine to Prevent Herpes Zoster and Postherpetic Neuralgia in Older Adults. N Engl J Med. 2005;352(22):2271–84.	1293
11	Ryan CJ, Smith MR, de Bono JS, Molina A, Logothetis CJ, de Souza P, et al. Abiraterone in metastatic prostate cancer without previous chemotherapy. N Engl J Med. 2013;368(2):138–48.	1288
12	Pfreundschuh M, Trümper L, Österborg A, Pettengell R, Trneny M, Imrie K, et al. CHOP-like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good-prognosis diffuse large-B-cell lymphoma: a randomised controlled trial by the MabThera International Trial (MInT) Group. Lancet Oncol. 2006;7(5):379–91.	1206
13	Nielsen N, Wetterslev J, Cronberg T, Erlinge D, Gasche Y, Hassager C, et al. Targeted Temperature Management at 33°C versus 36°C after Cardiac Arrest. N Engl J Med. 2013;369(23):2197–206.	1169
14	Aoyama H, Shirato H, Tago M, Nakagawa K, Toyoda T, Hatano K, et al. Stereotactic Radiosurgery Plus Whole-Brain Radiation Therapy vs Stereotactic Radiosurgery Alone for Treatment of Brain Metastases. JAMA. 2006;295(21):2483.	1112
15	Leon MB, Smith CR, Mack MJ, Makkar RR, Svensson LG, Kodali SK, et al. Transcatheter or Surgical Aortic-Valve Replacement in Intermediate-Risk Patients. N Engl J Med. 2016;374(17):1609–20.	1102
16	King TE, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, et al. A Phase 3 Trial of Pirfenidone in Patients with Idiopathic Pulmonary Fibrosis. N Engl J Med. 2014;370(22):2083–92.	1050
17	Look AHEAD Research Group TLAR, Wing RR, Bolin P, Brancati FL, Bray GA, Clark JM, et al. Cardiovascular effects of intensive lifestyle intervention in type 2 diabetes. N Engl J Med. 2013;369(2):145–54.	1021
18	Green JB, Bethel MA, Armstrong PW, Buse JB, Engel SS, Garg J, et al. Effect of Sitagliptin on Cardiovascular Outcomes in Type 2 Diabetes. N Engl J Med. 2015;373(3):232–42.	991
19	Cornely OA, Maertens J, Winston DJ, Perfect J, Ullmann AJ, Walsh TJ, et al. Posaconazole vs. Fluconazole or Itraconazole Prophylaxis in Patients with Neutropenia. N Engl J Med. 2007;356(4):348–59.	890
20	Hézode C, Forestier N, Dusheiko G, Ferenci P, Pol S, Goeser T, et al. Telaprevir and Peginterferon with or without Ribavirin for Chronic HCV Infection. N Engl J Med. 2009;360(18):1839–50.	821

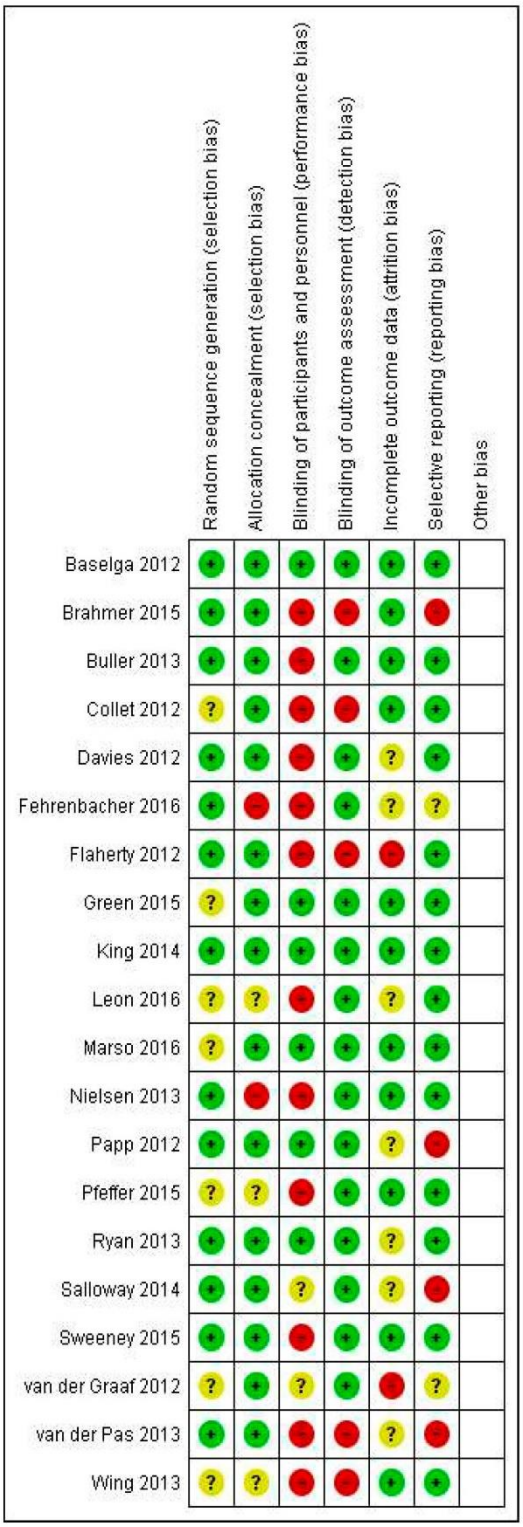
Tabla 4A (continuación). Listado de los estudios incluidos con sus correspondientes citas.

Núm.	Artículo	Total citas
21	Davies C, Pan H, Godwin J, Gray R, Arriagada R, Raina V, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. Lancet. 2013;381(9869):805–16.	799
22	Stopeck AT, Lipton A, Body J-J, Steger GG, Tonkin K, de Boer RH, et al. Denosumab compared with zoledronic acid for the treatment of bone metastases in patients with advanced breast cancer: a randomized, double-blind study. J Clin Oncol. 2010;28(35):5132–9.	757
23	van der Graaf WT, Blay J-Y, Chawla SP, Kim D-W, Bui-Nguyen B, Casali PG, et al. Pazopanib for metastatic soft-tissue sarcoma (PALETTE): a randomised, double-blind, placebo-controlled phase 3 trial. Lancet. 2012;379(9829):1879–86.	748
24	Janssen H LA, van Zonneveld M, Senturk H, Zeuzem S, Akarca US, Cakaloglu Y, et al. Pegylated interferon alfa-2b alone or in combination with lamivudine for HBeAg-positive chronic hepatitis B: a randomised trial. Lancet. 2005;365(9454):123–9.	742
25	Fehrenbacher L, Spira A, Ballinger M, Kowanetz M, Vansteenkiste J, Mazieres J, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. Lancet. 2016;387(10030):1837–46.	738
26	Salloway S, Sperling R, Fox NC, Blennow K, Klunk W, Raskind M, et al. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. N Engl J Med. 2014 Jan 23 [cited 2019 Jun 10];370(4):322–33.	685
27	Investigators TH-V. Edoxaban versus Warfarin for the Treatment of Symptomatic Venous Thromboembolism. N Engl J Med. 2013 Oct 10 [cited 2019 Jun 10];369(15):1406–15.	680
28	Albain KS, Swann RS, Rusch VW, Turrisi AT, Shepherd FA, Smith C, et al. Radiotherapy plus chemotherapy with or without surgical resection for stage III non-small-cell lung cancer: a phase III randomised controlled trial. Lancet. 2009 Aug 1 [cited 2019 Jun 11];374(9687):379–86.	679
29	Pfeffer MA, Claggett B, Diaz R, Dickstein K, Gerstein HC, Køber L V., et al. Lixisenatide in Patients with Type 2 Diabetes and Acute Coronary Syndrome. N Engl J Med. 2015 Dec 3 [cited 2019 Jun 10];373(23):2247–57.	670
30	Sweeney CJ, Chen Y-H, Carducci M, Liu G, Jarrard DF, Eisenberger M, et al. Chemohormonal Therapy in Metastatic Hormone-Sensitive Prostate Cancer. N Engl J Med. 2015;373(8):737–46.	655
31	Mayrand M-H, Duarte-Franco E, Rodrigues I, Walter SD, Hanley J, Ferenczy A, et al. Human Papillomavirus DNA versus Papanicolaou Screening Tests for Cervical Cancer. N Engl J Med. 2007;357(16):1579–88.	626
32	Bøtker HE, Kharbanda R, Schmidt MR, Bøttcher M, Kaltoft AK, Terkelsen CJ, et al. Remote ischaemic conditioning before hospital admission, as a complement to angioplasty, and effect on myocardial salvage in patients with acute myocardial infarction: a randomised trial. Lancet. 2010;375(9716):727–34.	593
33	Papp KA, Leonardi C, Menter A, Ortonne J-P, Krueger JG, Kricorian G, et al. Brodalumab, an Anti–Interleukin-17–Receptor Antibody for Psoriasis. N Engl J Med. 2012;366(13):1181–9.	581
34	Sachs GS, Nierenberg AA, Calabrese JR, Marangell LB, Wisniewski SR, Gyulai L, et al. Effectiveness of Adjunctive Antidepressant Treatment for Bipolar Depression. N Engl J Med. 2007;356(17):1711–22.	565
35	DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, et al. Cognitive Therapy vs Medications in the Treatment of Moderate to Severe Depression. Arch Gen Psychiatry. 2005 2019;62(4):409.	545
36	Nakamura H, Arakawa K, Itakura H, Kitabatake A, Goto Y, Toyota T, et al. Primary prevention of cardiovascular disease with pravastatin in Japan (MEGA Study): a prospective randomised controlled trial. Lancet. 2006;368(9542):1155–63.	539
37	The UK Standardisation of Breast Radiotherapy (START) Trial B of radiotherapy hypofractionation for treatment of early breast cancer: a randomised trial. Lancet. 2008 Mar 29 [cited 2019 Jun 11];371(9618):1098–107.	531
38	Collet J-P, Cuisset T, Rangé G, Cayla G, Elhadad S, Pouillot C, et al. Bedside Monitoring to Adjust Antiplatelet Therapy for Coronary Stenting. N Engl J Med [Internet]. 2012;367(22):2100–9.	531
39	van der Pas MH, Haglind E, Cuesta MA, Fürst A, Lacy AM, Hop WC, et al. Laparoscopic versus open surgery for rectal cancer (COLOR II): short-term outcomes of a randomised, phase 3 trial. Lancet Oncol. 2013;14(3):210–8.	527
40	Kupsch A, Benecke R, Müller J, Trottenberg T, Schneider G-H, Poewe W, et al. Pallidal Deep-Brain Stimulation in Primary Generalized or Segmental Dystonia. N Engl J Med. 2006;355(19):1978–90.	514

A. 2005-2010



B. 2012-2017

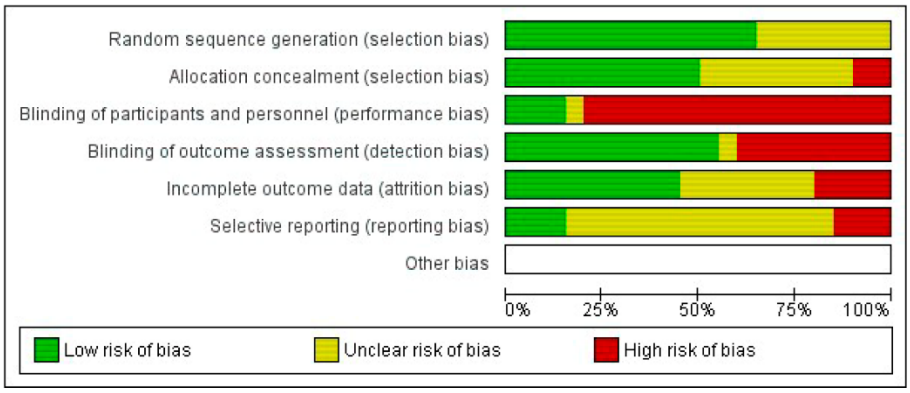


Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

Review Manager 5.3

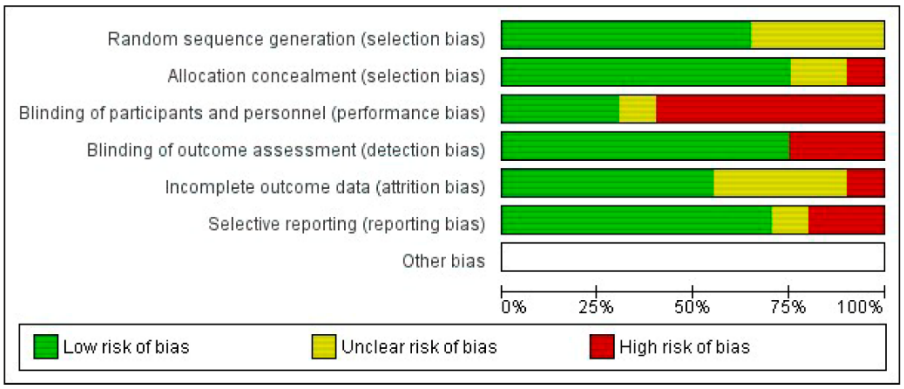
Figura 1A. Evaluación del riesgo de sesgo a nivel de estudio

A. Estudios 2005-2010



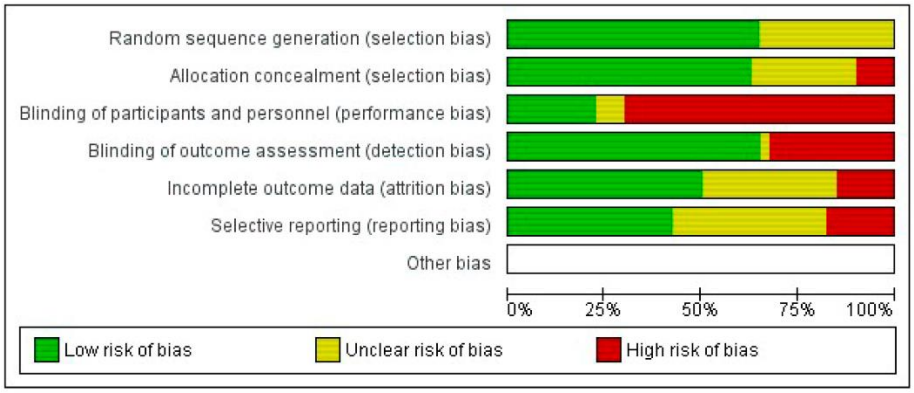
Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

B. Estudios 2012-2017



Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

C. Estudios 2005-2017



Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

Figura 2A. Resumen de la evaluación global del riesgo de sesgo