BMJ Open

Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019703
Article Type:	Research
Date Submitted by the Author:	20-Sep-2017
Complete List of Authors:	Page, Matthew; Monash University, School of Public Health and Preventive Medicine McKenzie, Joanne; Monash University, School of Public Health and Preventive Medicine Higgins, Julian; University of Bristol, Population Health Sciences, Bristol Medical School
<b>Primary Subject Heading</b>:	Research methods
Secondary Subject Heading:	Evidence based practice
Keywords:	Publication Bias, Bias (Epidemiology), Review Literature As Topic, Checklist

Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review

Matthew J Page[1,2], McKenzie JE[1], Higgins JPT[2]

1. School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
2. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

Correspondence to: Dr. Matthew Page, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne VIC 3004, Australia. Phone: +61 3 9903 0248. Email address: matthew.page@monash.edu

WORD COUNT: 3,591

ABSTRACT

OBJECTIVES: To examine the content and measurement properties of scales, checklists and domain-based tools that include a mechanism for assessing risk of reporting biases.

METHODS: In a systematic review, we searched for potentially relevant articles in Ovid MEDLINE, Ovid EMBASE, Ovid PsycINFO, and Google Scholar from inception to February 2017. One author screened all titles, abstracts and full text articles, and collected data on tool characteristics.

RESULTS: We identified 18 tools that include an assessment of the risk of reporting bias. The tools varied with regards to the type of reporting bias assessed (e.g. bias due to selective publication, bias due to selective non-reporting), and the level of assessment (e.g. for the study as a whole, a particular result within a study, or a particular synthesis of studies). Various criteria are used across tools to designate a synthesis as being at "high" risk of bias due to selective publication (e.g. evidence of funnel plot asymmetry, use of non-comprehensive searches). However, the relative weight assigned to each criterion in the overall judgement is not clear for most of these tools. Tools for assessing risk of bias due to selective non-reporting guide users to assess a study, or an outcome within a study, as "high" risk of bias if no results are reported for an outcome. However, assessing the corresponding risk of bias in a synthesis that is missing the non-reported outcomes is not within the scope of any of these tools. Inter-rater agreement estimates were available for five tools.

CONCLUSION: There are several limitations of existing tools for assessing risk of reporting biases, in terms of their scope, guidance for reaching risk of bias judgements, and measurement properties. Development and evaluation of a new, comprehensive tool, is required to try and overcome present limitations.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- Tools for assessing risk of reporting biases, and studies evaluating their measurement properties, were identified by searching several relevant databases using a search string developed in conjunction with an information specialist.
- Detailed information on the content and measurement properties of existing tools was collected, providing readers with pertinent information to help decide which tools to use in future evidence syntheses.
- Screening of articles and data collection were performed by one author only, so it is possible that some relevant articles were missed, or that errors in data collection were made.
- The search of grey literature was not comprehensive, so it is possible that there are other tools for assessing risk of reporting biases, and unpublished studies evaluating measurement properties, that were omitted from this review.

BACKGROUND

The credibility of evidence syntheses can be compromised by reporting biases, which arise when dissemination of research findings is influenced by the nature of the results[1]. For example, there may be bias due to selective publication, where a study is only published if the findings are considered interesting (also known as publication bias)[2]. In addition, bias due to selective non-reporting may occur, where findings (e.g. estimates of intervention efficacy or an association between exposure and outcome) that are statistically non-significant are not reported or are partially reported in a paper (e.g. stating only that "P>0.05")[3]. Alternatively, there may be bias in selection of the reported result, where authors perform multiple analyses for a particular outcome/association, yet only report the result which yielded the most favourable effect estimate[4]. Evidence from cohorts of clinical trials followed from inception suggest that biased dissemination is common. Specifically, on average, half of all trials are not published[1 5], trials with statistically significant results are twice as likely to be published[5], and a third of trials have outcomes that are omitted, added or modified between protocol and publication[6].

Audits of systematic review conduct suggest that most systematic reviewers do not assess risk of reporting biases[7-9]. For example, in a cross-sectional study of 300 systematic reviews indexed in MEDLINE® in February 2014[7], the risk of bias due to selective publication was not considered in 56% of reviews. A common reason for not doing so was that the small number of included studies, or inability to perform a meta-analysis, precluded the use of funnel plots. Only 19% of reviews included a search of a trial registry to identify completed but unpublished trials or pre-specified but non-reported outcomes, and only 7% included a search of another source of data disseminated outside of journal articles. The risk of bias due to selective non-reporting in the included studies was assessed in only 24% of reviews[7]. Another study showed that authors of Cochrane reviews routinely record whether any measured outcomes were not reported in the included trials, yet rarely consider if such non-reporting could have biased the results of a synthesis[10].

It is unclear why so few systematic reviewers assess risk of reporting biases adequately, since several tools (i.e. structured instruments such as scales, checklists, or domain-based tools) have been developed to assess these sources of bias. However, it is possible that existing tools are not fit for purpose. Previous researchers have summarised the characteristics of tools designed to assess various sources of bias in randomized trials[11-13], non-randomized studies of interventions (NRSI)[13 14], diagnostic test accuracy studies[15], and systematic reviews[13 16]. Others have summarised the performance of statistical methods developed to detect or adjust for reporting biases[17-19]. However, no prior review has focused specifically on tools for assessing the risk of reporting biases. Therefore, the aim of this research was to conduct a systematic review of the content and measurement properties of such tools.

METHODS

Protocol

Methods for this systematic review were pre-specified in a protocol, which was uploaded to the Open Science Framework in February 2017 (https://osf.io/9ea22/).

Eligibility criteria

Papers were included if the authors described a tool that was designed for use by individuals performing evidence syntheses to assess risk of reporting biases in the included studies or in their synthesis of studies. Tools could assess any type of reporting bias, including bias due to selective publication, bias due to selective non-reporting, or bias in selection of the reported result. Tools could assess the risk of reporting biases in any type of study (e.g. randomized trial of intervention, diagnostic test accuracy study, observational study estimating prevalence of an exposure), and in any type of result (e.g. estimate of intervention efficacy or harm, association between exposure and outcome, estimate of diagnostic accuracy). Eligible tools could take any form, including scales, checklists, and domain-based tools. To be considered a scale, each item had to have a numeric score attached to it, so that an overall summary score could be calculated[11]. To be considered a checklist, the tool had to include multiple questions, but the developers' intention was not to attach a numerical score to each response, or to calculate an overall score[12]. Domain-based tools require users to judge risk of bias or quality within specific domains, and to record the information on which each judgement is based[20].

Tools with a broad scope, for example, to assess multiple sources of bias or the overall quality of the body of evidence, were eligible if one of the items covered risk of reporting bias. Multi-dimensional tools with a statistical component were also eligible (e.g. those that require users to respond to a set of questions about the comprehensiveness of the search, as well as to perform statistical tests for funnel plot asymmetry). In addition, any studies that evaluated the measurement properties of existing tools (e.g. construct validity, inter-rater agreement, time taken to complete assessments) were eligible for inclusion. Papers were eligible regardless of the date or format of publication, but were limited to those written in English.

The following were ineligible:

- articles or book chapters providing guidance on how to address reporting biases, but which do not include a structured tool that can be applied by users (e.g. the 2011 Cochrane Handbook chapter on reporting biases[21]);
- tools developed or modified for use in one particular systematic review;
- tools designed to appraise published systematic reviews, such as the ROBIS tool[22] or AMSTAR[23];
- articles that focus on the development or evaluation of statistical methods to detect or adjust for reporting biases.

Search methods

On 9 February 2017, one author (MJP) searched for potentially relevant records in Ovid MEDLINE (January 1946 to February 2017), Ovid EMBASE (January 1980 to February 2017), and Ovid PsycINFO (January 1806 to February 2017). The search strategies included terms relating to reporting bias, which were combined with a search string used previously by Whiting et al. to identify risk of bias/quality assessment tools[16] (see full Boolean search strategies in online supplementary table S1).

To capture any tools not published by formal academic publishers, we searched Google Scholar using the phrase “reporting bias tool OR risk of bias”. One author (MJP) screened the titles of the first 300 records, as recommended by Haddaway et al.[24]. To capture any papers that may have been missed by all searches, one author (MJP) screened the references of included articles.

Study selection and data collection

One author (MJP) screened all titles and abstracts retrieved by the searches. The same author screened any full text articles retrieved. One author (MJP) collected data from included papers using a standardised data collection form. The following data on included tools were collected:

- type of tool (scale, checklist, or domain-based tool);
- types of reporting bias addressed by the tool;
- level of assessment (i.e. whether users direct assessments at the synthesis or at the individual studies included in the synthesis);
- whether the tool is designed for general use (generic) or targets specific study designs or topic areas (specific);
- items included in the tool;
- how items within the tool are rated;
- methods used to develop the tool (e.g. Delphi study, expert consensus meeting);
- availability of guidance to assist with completion of the tool (e.g. guidance manual).

The following data from studies evaluating measurement properties of an included tool were collected:

- tool evaluated;
- measurement properties evaluated (e.g. inter-rater agreement);
- number of syntheses/studies evaluated;
- publication year of syntheses/studies evaluated;
- areas of health care addressed by syntheses/studies evaluated;
- number of assessors;
- effect estimate and precision of measurement properties (e.g. weighted kappa).

Data analysis

We summarised the characteristics of included tools in tables. We calculated the median (interquartile range (IQR)) number of items across all tools, and tabulated the frequency of different criteria used in tools to denote a judgement of "high" risk of reporting bias. We summarised estimates of measurement properties, such as weighted kappa to estimate inter-rater agreement[25], by calculating the range of values across studies. For studies reporting weighted kappa, we categorised agreement according to the system proposed by Landis et al.[26], as poor (0.00), slight (0.01-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80), or almost perfect (0.81-1.00).

RESULTS

In total, 5,554 records were identified from the searches, of which we retrieved 165 for full text screening (Figure 1). The inclusion criteria were met by 42 reports summarising 18 tools (Table 1) and 17 studies evaluating the measurement properties of tools[3 4 20 27-65]. A list of excluded papers is presented in online supplementary Table S2.

Table 1. List of included tools

Article ID	Tool	Scope of tool	Types of reporting biases assessed			Level of assessment[a]
			Selective publication	Selective non-reporting	Selection of the reported result	
Balshem 2013[27]	AHRQ outcome and analysis reporting bias framework	Reporting bias only		✓	✓	Specific outcome/ result in a study
Berkman 2013[28]	AHRQ tool for evaluating the risk of reporting bias	Reporting bias only	✓	✓		Specific synthesis of studies
Downes 2016[29]	AXIS tool (Appraisal tool for Cross-Sectional Studies)	Multiple sources of bias		✓		Study
Downs 1998[30]	Downs-Black tool	Multiple sources of bias			✓	Study
Guyatt 2011[32-36]	GRADE	Multiple sources of bias	✓	✓		Specific synthesis of studies
Hayden 2013[37]	QUIPS (Quality In Prognosis Studies) tool	Multiple sources of bias		✓		Study
Higgins 2011[20 38 39]	Cochrane risk of bias tool for randomized trials	Multiple sources of bias		✓	✓	Study
Higgins 2016[40 41]	RoB 2.0 revised tool for assessing risk of bias in randomized trials	Multiple sources of bias			✓	Specific result in a study
Hoojimans 2014[42]	SYRCLE's RoB tool (SYstematic Review Centre for Laboratory animal Experimentation)	Multiple sources of bias		✓	✓	Study
Kim 2013[43]	RoBANS (Risk of Bias Assessment Tool for Nonrandomized Studies)	Multiple sources of bias		✓	✓	Study
Kirkham	ORBIT-I (Outcome Reporting Bias In Trials)	Reporting bias		✓		Specific outcome

Article ID	Tool	Scope of tool	Types of reporting biases assessed			Level of assessment[a]
			Selective publication	Selective non-reporting	Selection of the reported result	
2010[3 31]	classification system for benefit outcomes	only				in a study
Meader 2014[44 45]	SAQAT (Semi-Automated Quality Assessment Tool)	Multiple sources of bias	✓	✓		Specific synthesis of studies
Reid 2015[46]	Selective reporting bias algorithm	Reporting bias only		✓	✓	Study
Saini 2014[47]	ORBIT-II (Outcome Reporting Bias In Trials) classification system for harm outcomes	Reporting bias only		✓		Specific outcome/ result in a study
Salanti 2014[48 49]	Framework for evaluating the quality of evidence from a network meta-analysis	Multiple sources of bias	✓	✓		Specific synthesis of studies
Sterne 2016[4]	ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool	Multiple sources of bias			✓	Specific result in a study
Viswanathan 2012[50]	RTI Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures	Multiple sources of bias		✓		Study
Viswanathan 2013[51]	RTI Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures	Multiple sources of bias		✓		Study

[a]Level of assessment classified as: "study" when assessments are directed at a study as a whole (e.g. tool used to assess whether *any* outcomes in a study were not reported); "specific outcome/result in a study" when assessments are directed at a specific outcome or result within a study (e.g. tools used to assess whether a particular outcome, such as pain, was not reported) or; "specific synthesis of studies" when assessments are directed at a specific synthesis (e.g. tool used to assess whether a particular synthesis, such as a meta-analysis of pain, is missing unpublished studies).

General characteristics of included tools

Nearly all of the included tools (16/18 [89%]) were domain-based, where users judge risk of bias or quality within specific domains (Table 2; individual characteristics of each tool are presented in online supplementary Table S3). All tools were designed for generic rather than specific use. Five tools focused solely on the risk of reporting biases[3 27 28 46 47]; the remainder addressed reporting biases and other sources of bias/methodological quality (e.g. problems with randomization, lack of blinding). Half of the tools (9/18 [50%]) addressed only one type of reporting bias (e.g. bias due to selective non-reporting only).

The content of the included tools was informed by various sources of data. The most common included a literature review of items used in existing tools or a literature review of empirical evidence of bias (9/18 [50%]), ideas generated at an expert consensus meeting (8/18 [44%]) and pilot feedback on a preliminary version of the tool (7/18 [39%]). The most common type of guidance available for the tools was a brief annotation per item/response option (9/18 [50%]). A detailed guidance manual is available for four (22%) tools.

Table 2. Summary of general characteristics of included tools

Characteristic	**Summary data (n = 18 tools)**
Type of tool	
Domain-based	16 (89%)
Checklist	1 (6%)
Scale	1 (6%)
Scope of tool	
Assessment of reporting bias only	5 (28%)
Assessment of multiple sources of bias/quality	13 (72%)
Types of reporting bias assessed	
Bias due to selective publication only	0 (0%)
Bias due to selective non-reporting only	6 (33%)
Bias in selection of the reported result only	3 (17%)
Bias due to selective publication and bias due to selective non-reporting	4 (22%)
Bias due to selective non-reporting and bias in selection of the reported result	5 (28%)
Total number of items in the tool	7 (5-13)
Number of items relevant to risk of reporting bias	1 (1-2)
Number of response options for risk of reporting bias judgement	3 (3-3)
Types of study designs to which the tool applies	
Randomized trials only	5 (28%)
Systematic reviews only	3 (17%)
Non-randomized studies of interventions only	2 (11%)
Randomized trials and non-randomized studies of interventions	2 (11%)
Non-randomized studies of interventions or exposures	2 (11%)
Other (cross-sectional studies, animal studies, network meta-analyses, prognosis studies)	4 (22%)
Level of assessment of risk of reporting bias	
Study as a whole	9 (50%)
Specific outcome/result in a study	5 (28%)
Specific synthesis of studies	4 (22%)
Data sources used to inform tool content[a]	
Literature review (e.g. of items in existing tools, or empirical evidence)	9 (50%)
Ideas generated at expert consensus meeting	8 (44%)
Pilot feedback on preliminary version of the tool	7 (39%)

Characteristic	Summary data (n = 18 tools)
Psychometric or cognitive testing	5 (28%)
Other (e.g. adaptation of existing tool)	5 (28%)
Delphi study responses	2 (11%)
No methods stated	2 (11%)
Guidance available	
Brief annotation per item/response option	9 (50%)
Detailed guidance manual	4 (22%)
Worked example for each response option	2 (11%)
Detailed annotation per item/response option	1 (6%)
None	2 (11%)

Summary data given as number (percent) or median (IQR).
[a]The percentages in this category do not sum to 100% since the development of some tools was informed by multiple data sources.

Tool content

Four tools include items for assessing risk of bias due to both selective publication and selective non-reporting[28 32 44 48]. One of these tools (the AHRQ tool for evaluating the risk of reporting bias[28]) directs users to assess a particular synthesis, where a single risk of bias judgement is made based on information about unpublished studies and underreported outcomes. In the other three tools[32 44 48], the different sources of reporting bias are assessed in separate domains.

Five tools[20 27 42 43 46] guide users to assess risk of bias due to both selective non-reporting and selection of the reported result (that is, problems with outcomes/results that *are not* reported and those that *are* reported, respectively). Four of these tools, which include the Cochrane risk of bias tool for randomized trials[20] and three others which are based on the Cochrane tool[42 43 46], direct assessments at the study level. That is, a whole study is rated at “high” risk of reporting bias if *any* outcome/result in the study has been omitted, or fully reported, on the basis of the findings.

Some of the tools designed to assess the risk of bias due to selective non-reporting (e.g. ORBIT tools[3 47] and the AHRQ outcome reporting bias framework[27]) ask users to assess, for particular outcomes of interest, whether the outcome was not reported or only partially reported in the study on the basis of its results. This allows users to perform multiple outcome-level assessments of the risk of reporting bias (rather than one assessment for the study as a whole). However, assessing the corresponding risk of bias in a synthesis that is missing the non-reported outcome, is not within the scope of these tools.

A variety of criteria are used in existing tools to inform a judgement of “high” risk of bias due to selective publication (Table 3), selective non-reporting (Table 4), and selection of the reported result (Table 5) (more detail is provided in online supplementary Table S4). In the four tools with an assessment of risk of bias due to selective publication, “high” risk criteria include evidence of funnel plot asymmetry, discrepancies between published and unpublished studies, use of non-comprehensive searches, and presence of small, “positive” studies with for-profit interest (Table 3). However, not all of these criteria appear in all tools (only evidence of funnel plot asymmetry does), and the relative weight assigned to each criterion in the overall risk of reporting bias judgement is clear for only one tool (the Semi-Automated Quality Assessment Tool (SAQAT)[44 45]).

All 15 tools with an assessment of the risk of bias due to selective non-reporting suggest that the risk of bias is “high” when it is clear that an outcome was measured but no results were reported (Table 4). Fewer of these tools (n=8 [53%]) also recommend a “high” risk judgement when results for an outcome are partially reported (e.g. it is stated that the result was non-significant, but no effect estimate or summary statistics are presented).

The eight tools including an assessment of the risk of bias in selection of the reported result recommend various criteria for a “high” risk judgement (Table 5). These include when some outcomes that were not pre-specified are added post-hoc (in 4 [50%] tools), or when it is likely that the reported result for a particular outcome has been selected, on the basis of the findings, from amongst multiple outcome measurements or analyses within the outcome domain (in 2 [25%] tools).

Table 3. Criteria used in existing tools to inform a judgement of "high" risk of bias due to selective publication

"High" risk of bias criteria proposed in existing tools	**AHRQ RRB**	**GRADE**	**SAQAT**	**NMA-Quality**	**Total n (%)**
Assessment directed at a specific synthesis (e.g. meta-analysis)					
Evidence of funnel plot asymmetry (based on visual inspection of funnel plot or statistical test for funnel plot asymmetry)	✓	✓	✓	✓	4 (100)
Smaller studies tend to demonstrate more favourable results (based on visual assessment, without funnel plot)	✓				1 (25)
Clinical decision would differ for estimates from a fixed-effect versus a random-effects model, because the findings from a fixed-effect model are closer to the null	✓				1 (25)
Substantial heterogeneity in the meta-analysis cannot be explained by some clinical or methodological factor	✓				1 (25)
At least one study is affected by selective outcome reporting, selective analysis reporting, non-publication or non-accessibility	✓				1 (25)
Presence of small (often "positive") studies with for-profit interest in the synthesis		✓		✓	2 (50)
Presence of early studies (i.e. set of small, "positive" trials addressing a novel therapy) in the synthesis		✓		✓	2 (50)
Discrepancy in findings between published and unpublished trials		✓	✓	✓	3 (75)
Search strategies were not comprehensive		✓	✓	✓	3 (75)
Methods to identify all available evidence were not comprehensive		✓		✓	2 (50)
Grey literature were not searched			✓		1 (25)
Restrictions to study selection on the basis of language were applied			✓		1 (25)
Industry influence may apply to studies included in the synthesis			✓		1 (25)

AHRQ RRB = AHRQ tool for evaluating the risk of reporting bias[28]; GRADE = GRADE rating of quality of evidence[33-36]; NMA-Quality = Framework for evaluating the quality of evidence from a network meta-analysis[48]; SAQAT = Semi-Automated Quality Assessment Tool[44 45].

Table 4. Criteria used in existing tools to inform a judgement of "high" risk of bias due to selective non-reporting

"High" risk of bias criteria proposed in existing tools	AHRQ ORB	AHRQ RRB	AXIS	GRADE	QUIPS	RoB 1.0	SYRCLE RoB	RoBANS	ORBIT-I	SAQAT	Reid	ORBIT-II	NMA-Quality	RTI 2012	RTI 2013	Total n (%)
Assessment directed at study as a whole																
One or more outcomes of interest were clearly measured, but no results were reported			✓	✓	✓	✓	✓	✓		✓	✓		✓	✓	✓	11 (73)
One or more outcomes of interest are reported incompletely so that they cannot be entered in a meta-analysis				✓		✓		✓					✓			4 (27)
The study report fails to include results for a key outcome that would be expected to have been reported for such a study				✓		✓	✓	✓					✓	✓	✓	7 (47)
Assessment directed at a specific outcome																
Particular outcome clearly measured, but no results were reported	✓	✓							✓			✓				4 (27)
Particular outcome of interest is reported incompletely so that it cannot be entered in a meta-analysis (typically stating only that P>0.05).	✓	✓							✓			✓				4 (27)

"High" risk of bias criteria proposed in existing tools	AHRQ ORB	AHRQ RRB	AXIS	GRADE	QUIPS	RoB 1.0	SYRCLE RoB	RoBANS	ORBIT-I	SAQAT	Reid	ORBIT-II	NMA-Quality	RTI 2012	RTI 2013	Total n (%)
Clinical judgment says particular outcome is likely to have been measured and analysed but not reported on the basis of its results	✓	✓							✓			✓				4 (27)

AHRQ ORB = AHRQ outcome and analysis reporting bias framework[27]; AHRQ RRB = AHRQ tool for evaluating the risk of reporting bias[28]; AXIS = Appraisal tool for Cross-Sectional Studies[29]; GRADE = GRADE rating of quality of evidence[33-36]; NMA-Quality = Framework for evaluating the quality of evidence from a network meta-analysis[48]; ORBIT-I = Outcome Reporting Bias In Trials classification system for benefit outcomes[3 31]; ORBIT-II = Outcome Reporting Bias In Trials classification system for harm outcomes[47]; QUIPS = Quality In Prognosis Studies tool[37]; Reid = Reid et al. selective reporting bias algorithm[46]; RoB 1.0 = Cochrane risk of bias tool for randomized trials[20 38 39]; RoBANS = Risk of Bias Assessment Tool for Nonrandomized Studies[43]; RTI 2012 = RTI Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures[50]; RTI 2013 = RTI Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures[51]; SAQAT = Semi-Automated Quality Assessment Tool[44 45]; SYRCLE RoB = SYstematic Review Centre for Laboratory animal Experimentation risk of bias tool[42].

Table 5. Criteria used in existing tools to inform a judgement of "high" risk of bias in selection of the reported result

"High" risk of bias criteria proposed in existing tools	AHRQ ORB	Downs-Black	RoB 1.0	RoB 2.0	SYRCLE RoB	RoBANS	Reid	ROBINS-I	Total n (%)
Assessment directed at study as a whole									
One or more reported outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse event)			✓		✓	✓	✓		4 (50)
One or more outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified			✓		✓				2 (15)
One or more retrospective, unplanned, subgroup analysis was reported		✓							1 (13)
Any analyses that had not been planned at the outset of the study were not clearly indicated		✓							1 (13)
Assessment directed at a specific outcome/result									
Particular outcome was not pre-specified but results were reported	✓								1 (13)
Reported result for a particular outcome is likely to have been selected, on the basis of the findings, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain				✓				✓	2 (25)
Reported result for a particular outcome is likely to have been selected, on the basis of the findings, from multiple analyses of the data				✓				✓	2 (25)
Reported result for a particular outcome is likely to have been selected, on the basis of the findings, from different subgroups								✓	1 (13)

AHRQ ORB = AHRQ outcome and analysis reporting bias framework[27]; Downs-Black = Downs-Black tool[30]; Reid = Reid et al. selective reporting bias algorithm[46]; RoB 1.0 = Cochrane risk of bias tool for randomized trials[20 38 39]; RoB 2.0 = Revised tool for assessing risk of bias in randomized trials[40 41]; RoBANS = Risk of Bias Assessment Tool for Nonrandomized Studies[43]; ROBINS-I = Risk Of Bias In Non-randomized Studies of Interventions tool[4]; SYRCLE RoB = SYstematic Review Centre for Laboratory animal Experimentation risk of bias tool[42].

General characteristics of studies evaluating measurement properties of included tools

Despite identifying 17 studies that evaluated one of the included tools, data on measurement properties of the risk of reporting bias component were available for 12 studies only[42 43 53-59 61 63 65] (the other five studies include only data on properties of the multi-dimensional tool as a whole[30 52 60 62 64]) (online supplementary Table S5). Nearly all 12 studies (11 [92%]) evaluated inter-rater agreement between two assessors; eight of these studies reported weighted kappa (κ) values, but only two described the weighting scheme[54 61]. Eleven studies[42 43 53-59 63 65] evaluated the properties of tools for assessing the risk of bias in a study due to selective non-reporting or bias in selection of the reported result, in which a median of 40 (IQR 32-109) studies were assessed. One study[61] evaluated a tool for assessing the risk of bias in a synthesis due to selective publication, in which 44 syntheses were assessed. All studies involved two assessors.

Results of evaluation studies

Five studies[53 55-57 59] included data on the inter-rater agreement of assessments of risk of bias due to selective non-reporting using the Cochrane risk of bias tool for randomized trials[20] (Table 6). Weighted kappa (κ) values in four studies[53 55-57] ranged from 0.13 to 0.50, suggesting slight to moderate agreement[26]. In the other study[59], the percent agreement in selective non-reporting assessments in trials that were included in two different Cochrane reviews was low (43% of judgements were in agreement). Two other studies found that inter-rater agreement of selective non-reporting assessments were substantial for SYRCLE's RoB tool (κ = 0.62, n = 32)[42], but poor for the RoBANS tool (κ = 0, n = 39)[43]. There was substantial agreement between raters in the assessment of risk of bias due to selective publication using the SAQAT (κ = 0.63, n = 29)[61]. The inter-rater agreement of assessments of risk of bias in selection of the reported result using the ROBINS-I tool[4] was moderate for NRSI included in a review of the effect of cyclooxygenase-2 (COX-2) inhibitors on cardiovascular events, and substantial for NRSI included in a review of the effect of thiazolidinediones on cardiovascular events[54].

Table 6. Reported measurement properties of tools with an assessment of the risk of reporting bias

Study ID	Tool	Measurement property	Sample size	Areas of health care addressed	Weighted kappa (95% CI)	Weighting scheme	Interpretation of kappa[a]
Armijo-Olivo 2014[53]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between two external reviewers)	87	Musculoskeletal, cardiorespiratory, neurological, and gynaecological conditions	0.5 (CI not reported)	Not described	Moderate agreement
Armijo-Olivo 2014[53]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between two external reviewers and Cochrane reviewers)	87	See above	0.13 (CI not reported)	Not described	Slight agreement
Hartling 2009[55]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting	163	Child health	0.13 (95% CI -0.05 to 0.31)	Not described	Slight agreement
Hartling 2011[56]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting	107	Asthma	0.4 (95% CI 0.14 to 0.67)	Not described	Fair agreement
Hartling 2012[57] [58]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between two reviewers, all trials)	124	Varied	0.27 (95% CI 0.06 to 0.49)	Not described	Fair agreement
Hartling 2012[57] [58]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between pairs of reviewers across different centres, all trials)	30	Varied	0.08 (95% CI -0.09 to 0.26)	Not described	Slight agreement
Jordan 2017[59]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between judgements of trials appearing in two SRs)	28	Subfertility	Not reported[b]	Not applicable	Not applicable
Vale 2013[65]	RoB 1.0	Agreement between selective non-reporting assessments performed using published article only versus published article and data collected during the IPD process (i.e. trial protocol, data collection forms, IPD)	95	Cancer pain	Not reported[b]	Not applicable	Not applicable

Study ID	Tool	Measurement property	Sample size	Areas of health care addressed	Weighted kappa (95% CI)	Weighting scheme	Interpretation of kappa[a]
Hoojimans 2014[42]	SYRCLE RoB	Inter-rater agreement of assessments of risk of bias due to selective non-reporting	32	Animal studies (not specified)	0.62 (CI not reported)	Not described	Substantial agreement
Kim 2013[43]	RoBANS	Inter-rater agreement of assessments of risk of bias due to selective non-reporting	39	Depression, myocardial infarction, post-partum hemorrhage, chronic non-cancer pain	0 (CI not reported)	Not described	Poor agreement
Llewellyn 2015[61]	SAQAT	Inter-rater agreement of assessments of risk of bias due to selective publication (between two SAQAT raters)	29	Varied	0.63 (95% CI 0.17 to 1)	Quadratic	Substantial agreement
Llewellyn 2015[61]	SAQAT	Inter-rater agreement of assessments of risk of bias due to selective publication (between one rater using SAQAT and one using the standard GRADE approach)	15	Varied	Not reported[b]	Not applicable	Not applicable
Norris 2012[63]	ORBIT-I	Inter-rater agreement of ORBIT-I classifications of risk of bias due to selective non-reporting	40	Varied	Not calculated, as too little variation in judgements	Not applicable	Not applicable
Bilandzic 2016[54]	ROBINS-I	Inter-rater agreement of assessments of risk of bias in selection of the reported result	16	Thiazolidinediones and cardiovascular events	0.78 (CI not reported)	Linear	Substantial agreement
Bilandzic 2016[54]	ROBINS-I	Inter-rater agreement of assessments of risk of bias in selection of the reported result	21	COX-2 inhibitors and cardiovascular events	0.45 (CI not reported)	Linear	Moderate agreement

[a] Interpretation of kappa based on categorisation system defined by Landis et al.[26]. [b] Data presented as percent agreement, not weighted kappa. ORBIT-I = Outcome Reporting Bias In Trials classification system for benefit outcomes[3 31]; RoB 1.0 = Cochrane risk of bias tool for randomized trials[20 38 39]; RoBANS = Risk of Bias Assessment Tool for Nonrandomized Studies[43]; ROBINS-I = Risk Of Bias In Non-randomized Studies of Interventions tool[4]; SAQAT = Semi-Automated Quality Assessment Tool[44 45]; SRs = systematic reviews; SYRCLE RoB = SYstematic Review Centre for Laboratory animal Experimentation risk of bias tool[42].

DISCUSSION

From a systematic search of the literature, we identified 18 tools designed for use by individuals performing evidence syntheses to assess risk of reporting biases in the included studies or in their synthesis of studies. The tools varied with regard to the type of reporting bias assessed (e.g. bias due to selective publication, bias due to selective non-reporting), and the level of assessment (e.g. for the study as a whole, a particular outcome within a study, or a particular synthesis of studies). Various criteria are used across tools to designate a synthesis as being at "high" risk of bias due to selective publication (e.g. evidence of funnel plot asymmetry, use of non-comprehensive searches). However, the relative weight assigned to each criterion in the overall judgement is not clear for most of these tools. Tools for assessing risk of bias due to selective non-reporting guide users to assess a study, or an outcome within a study, as "high" risk of bias if no results are reported for an outcome. However, assessing the corresponding risk of bias in a synthesis that is missing the non-reported outcomes is not within the scope of any of these tools. Inter-rater agreement estimates were available for five tools[4 20 42 43 61], and ranged from poor to substantial; however the sample sizes of most evaluations were small, and few described the weighting scheme used to calculate kappa.

Strengths and limitations

There are several strengths of this research. Methods were conducted in accordance with a systematic review protocol (https://osf.io/9ea22/). Published articles were identified by searching several relevant databases using a search string developed in conjunction with an information specialist[16]. Detailed information on the content and measurement properties of existing tools was collected, providing readers with pertinent information to help decide which tools to use in future reviews. However, the findings need to be considered in light of some limitations. Screening of articles and data collection were performed by one author only. It is therefore possible that some relevant articles were missed, or that errors in data collection were made. The search for unpublished tools was not comprehensive (only Google Scholar was searched), so it is possible that other tools for assessing risk of reporting biases exist. Further, restricting the search to articles in English was done to expedite the review process, but may have resulted in loss of information about tools written in other languages, and additional evidence on measurement properties of tools.

Comparison with other studies

Other systematic reviews of risk of bias tools[11-16] have restricted inclusion to tools developed for particular study designs (e.g. randomized trials, diagnostic test accuracy studies), where the authors recorded all the sources of bias addressed. A different approach was taken in the current review, where all tools (regardless of study design) that address a particular source of bias were examined. By focusing on one source of bias only, the analysis of included items and criteria for risk of bias judgements was more detailed than that recorded previously. Some of the existing reviews of tools[14] considered tools that were developed or modified in the context of a specific systematic review. However, such tools were excluded from the current review as they are unlikely to have been developed systematically[14 66], and are difficult to find (all systematic reviews conducted during a particular period would need to have been examined for the search to be considered exhaustive).

Explanations and implications

Of the 18 tools identified, only four (22%) included a mechanism for assessing risk of bias due to selective publication, which is the type of reporting bias that has been investigated most often[2]. This is perhaps unsurprising given that hundreds of statistical methods to “detect” or “adjust” for bias due to selective publication have been developed[17]. These statistical methods may be considered by methodologists and systematic reviewers as the tools of choice for assessing this type of bias. However, there are many limitations of these statistical approaches, in terms of their underlying assumptions, and the software required to apply them, which has limited their use in practice[18 67]. As a result, a large number of systematic reviews currently ignore the risk of bias due to selective publication[7-9 68].

Our analysis suggests that the factors that need to be considered to assess risk of reporting biases adequately (e.g. comprehensiveness of the search, amount of data missing from the synthesis due to unpublished studies and underreported outcomes) are fragmented. A similar problem was occurring a decade ago with the assessment of risk of bias in randomized trials. Some authors assessed only problems with randomization, while others focused on whether trials were not "double blinded", or had any missing participant data[69]. It was not until all the important bias domains were brought together into a structured, domain-based tool to assess the risk of bias in randomized trials[20], that systematic reviewers started to consider risk of bias in trials comprehensively.

A similar initiative to link all the components needed to judge the risk of reporting biases into a comprehensive new tool may improve the credibility of evidence syntheses. In particular, there is an emergent need for a new tool to assess the risk that a synthesis (rather than individual studies) is affected by reporting biases. This tool could guide users to consider risk of bias due to both selective publication and selective non-reporting, given that both practices lead to the same consequence: results missing from the synthesis[10]. Careful thought would need to be given as to how to weigh up various pieces of information underpinning the judgement. For example, users will need guidance on how evidence of known, unpublished studies (as identified from trial registries or regulatory documents) should be considered alongside evidence that is more speculative (e.g. funnel plots suggesting that studies may be missing). Preparation of a detailed guidance manual may enhance the usability of the tool, and minimise misinterpretation and errors in assessments. Once developed, evaluations of the measurement properties of the tool, such as inter-rater agreement and construct validity, should be conducted to explore whether modifications to the tool are necessary.

Conclusions

There are several limitations of existing tools for assessing risk of reporting biases in studies or syntheses of studies, in terms of their scope, guidance for reaching risk of bias judgements, and measurement properties. Development and evaluation of a new, comprehensive tool, is required to try and overcome present limitations.

Acknowledgments

Not applicable.

Competing Interests

We have read the journal's policy and have the following competing interests: JPTH led or participated in the development of four of the included tools (the current Cochrane risk of bias tool for randomized trials, the RoB 2.0 tool for assessing risk of bias in randomized trials, the ROBINS-I tool for assessing risk of bias in non-randomized studies of interventions, and the framework for assessing quality of evidence from a network meta-analysis). MJP participated in the development of one of the included tools (the RoB 2.0 tool for assessing risk of bias in randomized trials). All authors are participating in the development of a new tool for assessing risk of reporting biases in systematic reviews.

Funding

There was no direct funding for this study. MJP is supported by an Australian National Health and Medical Research Council (NHMRC) Early Career Fellowship (1088535). JEM is supported by a NHMRC Australian Public Health Fellowship (1072366). JPTH is funded in part by Cancer Research UK Programme Grant C18281/A19169; is a member of the MRC Integrative Epidemiology Unit at the

University of Bristol, which is supported by the UK Medical Research Council and the University of Bristol (grant MC_UU_12013/9); and is a member of the MRC ConDuCT-II Hub (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures; grant MR/K025643/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

MJP conceived and designed the study, collected data, analysed the data, and wrote the first draft of the article. JM and JPTH provided input on the study design and contributed to revisions of the article. All authors approved the final version of the submitted article.

Data sharing statement

The study protocol, data collection form, and the raw data and statistical analysis code for this study are available on the Open Science Framework: https://osf.io/3jdaa/

References

1. Chan A-W, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *The Lancet* 2014;383(9913):257-66.

2. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010;14:8.

3. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.

4. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.

5. Schmucker C, Schell LK, Portalupi S, et al. Extent of non-publication in cohorts of studies approved by research ethics committees or included in trial registries. *PLoS One* 2014;9(12):e114023.

6. Jones CW, Keil LG, Holland WC, et al. Comparison of registered and published outcomes in randomized controlled trials: a systematic review. *BMC Med* 2015;13:282.

7. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med* 2016;13(5):e1002028.

8. Koletsi D, Valla K, Fleming PS, et al. Assessment of publication bias required improvement in oral health systematic reviews. *J Clin Epidemiol* 2016;76:118-24

9. Hedin RJ, Umberham BA, Detweiler BN, et al. Publication Bias and Nonreporting Found in Majority of Systematic Reviews and Meta-analyses in Anesthesiology Journals. *Anesth Analg* 2016;123(4):1018-25.

10. Page MJ, Higgins JPT. Rethinking the assessment of risk of bias due to selective reporting: a cross-sectional study. *Systematic reviews* 2016;5(1):108.

11. Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;16(1):62-73.

12. Armijo Olivo S, Macedo LG, Gadotti IC, et al. Scales to assess the quality of randomized controlled trials: a systematic review. *Phys Ther* 2008;88(2):156-75.

13. Bai A, Shukla VK, Bak G, et al. Quality Assessment Tools Project Report. Ottawa: Canadian Agency for Drugs and Technologies in Health, 2012.

14. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007;36(3):666-76.

15. Whiting P, Rutjes AW, Dinnes J, et al. A systematic review finds that diagnostic reviews fail to incorporate quality despite available tools. *J Clin Epidemiol* 2005;58(1):1-12.

16. Whiting P, Davies P, Savovic J, et al. Evidence to inform the development of ROBIS, a new tool to assess the risk of bias in systematic reviews, September 2013. Available from https://www.researchgate.net/publication/303312018_Evidence_to_inform_the_development_of_ROBIS_a_new_tool_to_assess_the_risk_of_bias_in_systematic_reviews [accessed 1 August 2017].

17. Mueller KF, Meerpohl JJ, Briel M, et al. Methods for detecting, quantifying and adjusting for dissemination bias in meta-analysis are described. *J Clin Epidemiol* 2016;80:25-33.

18. Jin ZC, Zhou XH, He J. Statistical methods for dealing with publication bias in meta-analysis. *Stat Med* 2015;34(2):343-60.

19. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002.

20. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

21. Sterne JAC, Egger M, Moher D. Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions Version 510 [updated March 2011] 2011.

22. Whiting P, Savovic J, Higgins JP, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;69:225-34.

23. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.

24. Haddaway NR, Collins AM, Coughlin D, et al. The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. *PLoS One* 2015;10(9):e0138237.

25. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20(1):37-46.

26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-74.

27. Balshem H, Stevens A, Ansari M, et al. Finding grey literature evidence and assessing for outcome and analysis reporting biases when comparing medical interventions: AHRQ and the Effective Health Care Program. (Prepared by the Oregon Health and Science University and the University of Ottawa Evidence-based Practice Centers under Contract Nos. 290-2007-10057-I and 290-2007-10059-I.) AHRQ Publication No. 13(14)-EHC096-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

28. Berkman ND, Lohr KN, Ansari M, et al. Chapter 15 Appendix A: A Tool for Evaluating the Risk of Reporting Bias (in Chapter 15: Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update). Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I). AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm

29. Downes MJ, Brennan ML, Williams HC, et al. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ open* 2016;6:e011458.

30. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.

31. Dwan K, Gamble C, Kolamunnage-Dona R, et al. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials* 2010;11:52.

32. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.

33. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol* 2011;64(4):407-15.

34. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64(12):1277-82.

35. Schünemann H, Brożek J, Guyatt G, et al. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. [Updated October 2013]. Available from http://gdt.guidelinedevelopment.org/app/handbook/handbook.html.

36. Santesso N, Carrasco-Labra A, Langendam M, et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. *J Clin Epidemiol* 2016

37. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.

38. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons 2008:187-241.

39. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org/.

40. Higgins JPT, Savović J, Page MJ, et al. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0), Version 20 October 2016. Available from http://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/rob2-0/ [accessed 19 September 2017].

41. Higgins JPT, Sterne JAC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials. *Cochrane Methods Cochrane Database of Systematic Reviews* 2016;10(Suppl 1):29-31.

42. Hooijmans CR, Rovers MM, de Vries RBM, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.

43. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013;66(4):408-14.

44. Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic reviews* 2014;3(1):82.

45. Stewart GB, Higgins JP, Schunemann H, et al. The use of Bayesian networks to assess the quality of evidence from research synthesis: 1. *PLoS One* 2015;10(3):e0114497.

46. Reid EK, Tejani AM, Huan LN, et al. Managing the incidence of selective reporting bias: a survey of Cochrane review groups. *Systematic reviews* 2015;4:85.

47. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ* 2014;349:g6501.

48. Salanti G, Giovane CD, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9(7):e99682.

49. Higgins JP, Del Giovane C, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *Value Health* 2014;17(7):A324.

50. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* 2012;65(2):163-78.

51. Viswanathan M, Berkman ND, Dryden DM, et al. AHRQ Methods for Effective Health Care. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or

Exposures: Further Development of the RTI Item Bank. Rockville (MD): Agency for Healthcare Research and Quality (US) 2013.

52. Armijo-Olivo S, Stiles CR, Hagen NA, et al. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract* 2012;18(1):12-8.

53. Armijo-Olivo S, Ospina M, da Costa BR, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. *PLoS One* 2014;9(5):e96920.

54. Bilandzic A, Fitzpatrick T, Rosella L, et al. Risk of Bias in Systematic Reviews of Non-Randomized Studies of Adverse Cardiovascular Effects of Thiazolidinediones and Cyclooxygenase-2 Inhibitors: Application of a New Cochrane Risk of Bias Tool. *PLoS Med* 2016;13(4):e1001987.

55. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.

56. Hartling L, Bond K, Vandermeer B, et al. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. *PLoS One* 2011;6(2):e17242.

57. Hartling L, Hamm M, Milne A, et al. AHRQ Methods for Effective Health Care. Validity and Inter-Rater Reliability Testing of Quality Assessment Instruments. Rockville (MD): Agency for Healthcare Research and Quality (US) 2012.

58. Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol* 2013;66(9):973-81.

59. Jordan VM, Lensen SF, Farquhar CM. There were large discrepancies in risk of bias tool judgments when a randomized controlled trial appeared in more than one systematic review. *J Clin Epidemiol* 2017;81:72-76.

60. Kumar A, Miladinovic B, Guyatt GH, et al. GRADE guidelines system is reproducible when instructions are clearly operationalized even among the guidelines panel members with limited experience with GRADE. *J Clin Epidemiol* 2016;75:115-8.

61. Llewellyn A, Whittington C, Stewart G, et al. The Use of Bayesian Networks to Assess the Quality of Evidence from Research Synthesis: 2. Inter-Rater Reliability and Comparison with Standard GRADE Assessment. *PLoS One* 2015;10(12):e0123511.

62. Mustafa RA, Santesso N, Brozek J, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol* 2013;66(7):736-42; quiz 42.e1-5.

63. Norris SL, Holmer HK, Ogden LA, et al. AHRQ Methods for Effective Health Care. Selective Outcome Reporting as a Source of Bias in Reviews of Comparative Effectiveness. Rockville (MD): Agency for Healthcare Research and Quality (US) 2012.

64. O'Connor SR, Tully MA, Ryan B, et al. Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. *BMC Res Notes* 2015;8:224.

65. Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. *BMJ* 2013;346:f1798.

66. Whiting PF, Rutjes AW, Westwood ME, et al. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol* 2013;66(10):1093-104.

67. Sterne JAC, Egger M, Moher D, et al. Chapter 10: Addressing reporting biases. In: Higgins JPT, Churchill R, Chandler J, et al., eds. Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0 (updated June 2017). Available from www.training.cochrane.org/handbook: Cochrane 2017.

68. Atakpo P, Vassar M. Publication bias in dermatology systematic reviews and meta-analyses. *J Dermatol Sci* 2016;82(2):69-74.

69. Lundh A, Gotzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. *BMC Med Res Methodol* 2008;8:22.

Figure legends

Figure 1. Flow diagram of identification, screening and inclusion of studies

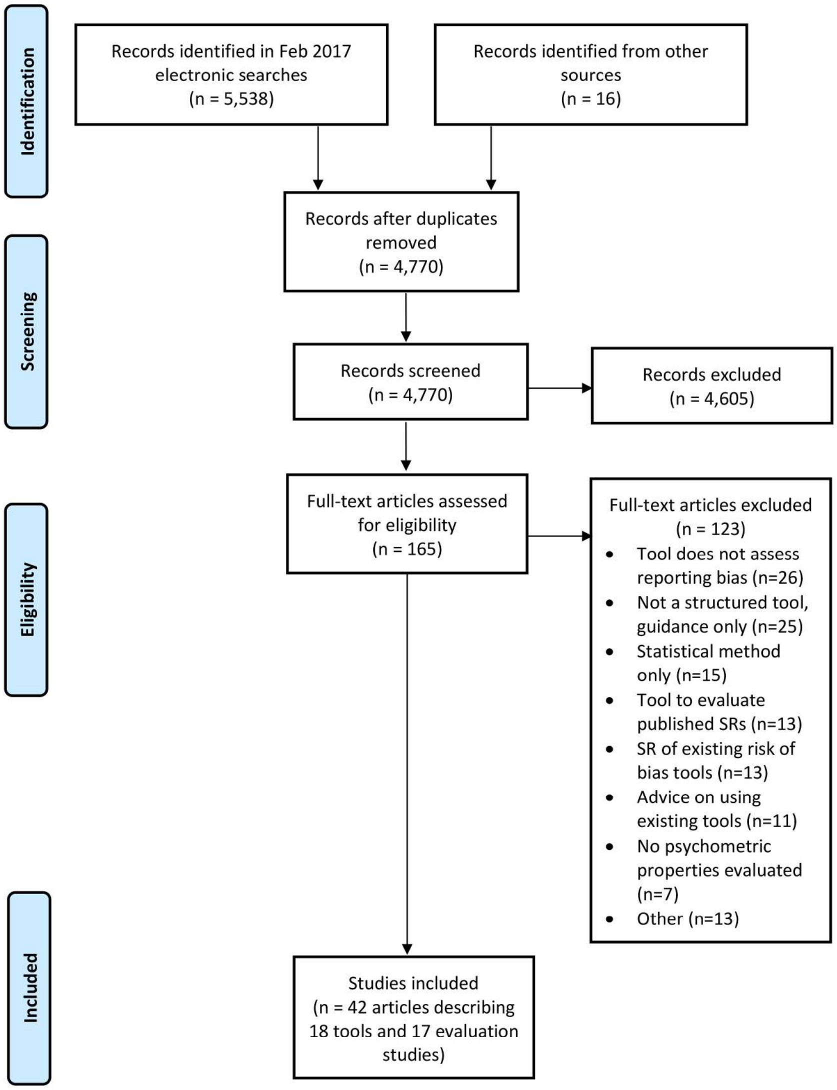
Identification
Records identified in Feb 2017 electronic searches (n = 5,538)
Records identified from other sources (n = 16)
Screening
Records after duplicates removed (n = 4,770)
Records screened (n = 4,770)
Records excluded (n = 4,605)
Eligibility
Full-text articles assessed for eligibility (n = 165)
Full-text articles excluded (n = 123)
• Tool does not assess reporting bias (n=26)
• Not a structured tool, guidance only (n=25)
• Statistical method only (n=15)
• Tool to evaluate published SRs (n=13)
• SR of existing risk of bias tools (n=13)
• Advice on using existing tools (n=11)
• No psychometric properties evaluated (n=7)
• Other (n=13)
Included
Studies included (n = 42 articles describing 18 tools and 17 evaluation studies)

Figure 1. Flow diagram of identification, screening and inclusion of studies

171x238mm (300 x 300 DPI)

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Table S1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7-8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	NA
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	NA

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	NA
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8, Fig 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	12
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	NA
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Table S3 and S4
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	NA
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	13-22
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	23
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	23-24
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	26
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	26-27

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.

Table S1. Search strategies

Database: Ovid MEDLINE(R) <1946 to 9 February 2017>
Search Strategy:

----

1 ((tool or tools or instrument$ or checklist$ or check list$ or scale or scales) and (quality or methodolog$ or method or methods)).ti.
2 (quality adj10 (score or scores or scoring or rating or rate) adj5 (methodolog$ or method or methods)).tw.
3 (guideline$ and (quality or methodolog$ or method or methods)).ti.
4 ((assess$ or apprais$ or critical$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).ti.
5 ((score or scores or scoring or rating or rate) and (quality or methodolog$ or method or methods)).ti.
6 ((quality or methodology) adj3 (review or meta-analys$ or metaanalys$) adj3 (assess$ or method$)).tw.
7 (quality adj3 article$).tw.
8 (critical$ adj2 (apprais$ or evaluat$)).tw.
9 ((apprais$ or evaluat$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
10 (guideline$ adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
11 or/1-10
12 Checklist/
13 11 or 12
14 Publication Bias/
15 exp "bias (epidemiology)"/
16 (bias adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
17 ((quality or bias or methodolog$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
18 (bias$ adj3 (publication$ or disseminat$ or language$ or reporting or grey or gray or citation$ or time delay or time lag or conference or abstract)).tw.
19 or/14-18
20 13 and 19

Database: Embase <1980 to 2017 Week 06>
Search Strategy:

----

1 "Review Literature as Topic"/
2 "meta analysis (topic)"/
3 meta analysis/
4 "systematic review (topic)"/
5 systematic review/
6 systematic review$.tw.
7 (meta-analys$ or metaanalys$).tw.
8 or/1-7
9 (bias adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
10 ((quality or bias or methodolog$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
11 (bias$ adj3 (publication$ or disseminat$ or language$ or reporting or grey or gray or citation$ or time delay or time lag or conference or abstract)).tw.

```
12   "internal validity"/
13   publishing/
14   or/9-13
15   ((tool or tools or instrument$ or checklist$ or check list$ or scale or scales) and (quality or
     methodolog$ or method or methods)).ti.
16   (quality adj10 (score or scores or scoring or rating or rate) adj5 (methodolog$ or method or
     methods)).tw.
17   (guideline$ and (quality or methodolog$ or method or methods)).ti.
18   ((assess$ or apprais$ or critical$) adj3 (systematic review$ or meta-analys$ or
     metaanalys$)).ti.
19   ((score or scores or scoring or rating or rate) and (quality or methodolog$ or method or
     methods)).ti.
20   ((quality or methodology) adj3 (review or meta-analys$ or metaanalys$) adj3 (assess$ or
     method$)).tw.
21   (quality adj3 article$).tw.
22   (critical$ adj2 (apprais$ or evaluat$)).tw.
23   ((apprais$ or evaluat$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
24   (guideline$ adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
25   or/15-24
26   checklist/
27   25 or 26
28   8 and 14 and 27
29   limit 28 to embase



Database: PsycINFO <1806 to February Week 1 2017>
Search Strategy:
--------------------------------------------------------------------------------
1    meta-analysis/
2    systematic review$.tw.
3    (meta-analys$ or metaanalys$).tw.
4    or/1-3
5    ((tool or tools or instrument$ or checklist$ or check list$ or scale or scales) and (quality or
     methodolog$ or method or methods)).ti.
6    (quality adj10 (score or scores or scoring or rating or rate) adj5 (methodolog$ or method or
     methods)).tw.
7    (guideline$ and (quality or methodolog$ or method or methods)).ti.
8    ((assess$ or apprais$ or critical$) adj3 (systematic review$ or meta-analys$ or
     metaanalys$)).ti.
9    ((score or scores or scoring or rating or rate) and (quality or methodolog$ or method or
1.  methods)).ti.
10   ((quality or methodology) adj3 (review or meta-analys$ or metaanalys$) adj3 (assess$ or
     method$)).tw.
11   (quality adj3 article$).tw.
12   (critical$ adj2 (apprais$ or evaluat$)).tw.
13   ((apprais$ or evaluat$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
14   (guideline$ adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
15   checklist/
16   or/5-15
17   (bias adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
```

18	((quality or bias or methodolog$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
19	(bias$ adj3 (publication$ or disseminat$ or language$ or reporting or grey or gray or citation$ or time delay or time lag or conference or abstract)).tw.
20	bias.mp.
21	or/17-20
22	4 and 16 and 21

Table S2. Excluded studies

Reference	**Reason for exclusion**
Armijo-Olivo S, Cummings GG, Fuentes J, Saltaji H, Ha C, Chisholm A, et al. Identifying items to assess methodological quality in physical therapy trials: a factor analysis. Physical Therapy 2014;94(9):1272-84.	Paper does not report on a structured tool
Armijo-Olivo S, Fuentes J, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: a descriptive analysis. BMC Medical Research Methodology 2013;13:116.	Systematic review of tools
Armijo-Olivo S, Fuentes J, Rogers T, Hartling L, Saltaji H, Cummings GG. How should we evaluate the risk of bias of physical therapy trials?: a psychometric and meta-epidemiological approach towards developing guidelines for the design, conduct, and reporting of RCTs in Physical Therapy (PT) area: a study protocol. Syst Rev 2013;2:88.	Protocol for development of new tool
Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. International Journal of Evidence-Based Healthcare 2015;13(3):132-40.	Refers to a tool to assess quality of published systematic reviews
Arrive L, Renard R, Carrat F, Belkacem A, Dahan H, Le Hir P, et al. A scale of methodological quality for clinical studies of radiologic examinations. Radiology 2000;217(1):69-74.	Tool does not assess reporting bias
Atakpo P, Vassar M. Publication bias in dermatology systematic reviews and meta-analyses. Journal of Dermatological Science 2016;82(2):69-74.	Describes statistical methods only
Ballard M, Montgomery P. Risk of bias in overviews of reviews: a scoping review of methodological guidance and four-item checklist. Research Synthesis Methods 2017;8(1):92-108.	Refers to a tool to assess quality of published systematic reviews
Balzer K. Assessing the quality of research needs to go beyond scoring: Commentary on Crowe and Sheppard (2011). International Journal of Nursing Studies 2012;49(8):1048-50.	Commentary
Bartlett WA, Braga F, Carobene A, Coskun A, Prusa R, Fernandez-Calle P, et al. A checklist for critical appraisal of studies of biological variation. Clinical Chemistry and Laboratory Medicine 2015;53(6):879-85.	Tool does not assess reporting bias
Bashir R, Dunn AG. Systematic review protocol assessing the processes for linking clinical trial registries and their published results. BMJ Open 2016;6(10):e013048.	Paper does not report on a structured tool
Beck NB, Becker RA, Boobis A, Fergusson D, Fowle JR, Goodman J, et al. Instruments for assessing risk of bias and other methodological criteria of animal studies: omission of well-established methods. Environmental Health Perspectives 2014;122(3):A66-7.	Commentary
Berkman ND, Lohr KN, Morgan LC, Kuo T-M, Morton SC. Interrater	Tool does not assess

Reference	Reason for exclusion
reliability of grading strength of evidence varies with the complexity of the evidence in systematic reviews. Journal of Clinical Epidemiology 2013;66(10):1105-17.e1.	reporting bias
Burda BU, Holmer HK, Norris SL. Limitations of A Measurement Tool to Assess Systematic Reviews (AMSTAR) and suggestions for improvement. Systematic Reviews 2016;5:58.	Refers to a tool to assess quality of published systematic reviews
Cartes-Velasquez RA, Manterola C, Aravena P, Moraga J. Reliability and validity of MINCIR scale for methodological quality in dental therapy research. Brazilian Oral Research 2014;28.	Tool does not assess reporting bias
Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Research Synthesis Methods 2012;3(2):161-76.	Describes statistical methods only
da Costa BR, Hilfiker R, Egger M. PEDro's bias: summary quality scores should not be used in meta-analysis. Journal of Clinical Epidemiology 2013;66(1):75-7.	Commentary
Dahm P. Raising the bar for systematic reviews with Assessment of Multiple Systematic Reviews (AMSTAR). BJU International 2017;119(2):193.	Refers to a tool to assess quality of published systematic reviews
Dalton DR, Aguinis H, Dalton CM, Bosco FA, Pierce CA. Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. Personnel Psychology 2012;65(2):221-49.	Paper does not report on a structured tool
David SP, Ware JJ, Chu IM, Loftus PD, Fusar-Poli P, Radua J, et al. Potential reporting bias in fMRI studies of the brain. PloS One 2013;8(7):e70104.	Paper does not report on a structured tool
Davino-Ramaya C, Krause LK, Robbins CW, Harris JS, Koster M, Chan W, et al. Transparency matters: Kaiser Permanente's National Guideline Program methodological processes. The Permanente Journal 2012;16(1):55-62.	Refers to a tool to assess quality of published systematic reviews
Dawson A, Raphael KG, Glaros A, Axelsson S, Arima T, Ernberg M, et al. Development of a quality-assessment tool for experimental bruxism studies: reliability and validity. Journal of Orofacial Pain 2013;27(2):111-22.	Tool does not assess reporting bias
Deshpande S, Misso K, Westwood M, Stirk L, De Kock S, Clayton D, et al. Not all cochrane reviews are good quality systematic reviews. Value in Health 2016;19(7):A371.	Refers to a tool to assess quality of published systematic reviews
Disher T, Benoit B, Johnston C, Campbell-Yeo M. Skin-to-skin contact for procedural pain in neonates: acceptability of novel systematic review synthesis methods and GRADEing of the evidence. Journal of Advanced Nursing 2017;73(2):504-19.	Paper does not report on a structured tool
Dreier M, Borutta B, Stahmeyer J, Krauth C, Walter U. Comparison of tools for assessing the methodological quality of primary and secondary studies in health technology assessment reports in Germany. GMS Health Technology Assessment 2010;6.	Systematic review of tools

Reference	Reason for exclusion
Dreyer N, Velentgas P, Duddy A, Westrich KD, Dubois RW. Grace checklist: Rating the strength of evidence for observational studies of comparative effectiveness. Value in Health 2012;15(4):A5.	Tool does not assess reporting bias
Dreyer NA, Velentgas P, Westrich K, Dubois R. The GRACE checklist for rating the quality of observational studies of comparative effectiveness: a tale of hope and caution. Journal of Managed Care & Specialty Pharmacy 2014;20(3):301-8.	Tool does not assess reporting bias
Dreyer NA, Velentgas P, Westrich K, Dubois RW. GRACE: A validated checklist for identifying robust observational studies of comparative effectiveness. Pharmacoepidemiol Drug Saf 2013;22:356.	Tool does not assess reporting bias
Dreyer NA, Velentgas P, Westrich KD, Dubois RW. There but for grace? a validated screening tool for quality observational studies of comparative effectiveness. Value in Health 2013;16(3):A21.	Tool does not assess reporting bias
Drucker AM, Fleming P, Chan A-W. Research Techniques Made Simple: Assessing Risk of Bias in Systematic Reviews. The Journal of Investigative Dermatology 2016;136(11):e109-e14.	Guidance on using existing tools
Dwan K, Altman DG, Clarke M, Gamble C, Higgins JP, Sterne JA, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. PLoS Med 2014;11(6):e1001666.	Paper does not report on a structured tool
Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. PLoS One 2013;8(7):e66844.	Paper does not report on a structured tool
Dwan K, Kirkham JJ, Williamson PR, Gamble C. Selective reporting of outcomes in randomised controlled trials in systematic reviews of cystic fibrosis. BMJ Open 2013;3(6).	No psychometric properties assessed
Fantony JJ, Gopalakrishna A, Noord MV, Inman BA. Reporting Bias Leading to Discordant Venous Thromboembolism Rates in the United States Versus Non-US Countries Following Radical Cystectomy: A Systematic Review and Meta-analysis. European Urology Focus 2016;2(2):189-96.	Paper does not report on a structured tool
Fitzgerald A, Coop C. Validation and modification of the Graphical Appraisal Tool for Epidemiology (GATE) for appraising systematic reviews in evidence-based guideline development. Health Outcomes Research in Medicine 2011;2(1):e51-e9.	Refers to a tool to assess quality of published systematic reviews
Frosi G, Riley RD, Williamson PR, Kirkham JJ. Multivariate meta-analysis helps examine the impact of outcome reporting bias in Cochrane rheumatoid arthritis reviews. J Clin Epidemiol 2015;68(5):542-50.	No psychometric properties assessed
Furukawa TA, Miura T, Chaimani A, Leucht S, Cipriani A, Noma H, et al. Using the contribution matrix to evaluate complex study limitations in a network meta-analysis: a case study of bipolar maintenance pharmacotherapy review. BMC Res Notes 2016;9:218.	Describes statistical methods only

Reference	Reason for exclusion
Ghogomu EAT, Maxwell LJ, Buchbinder R, Rader T, Pardo Pardo J, Johnston RV, et al. Updated method guidelines for cochrane musculoskeletal group systematic reviews and metaanalyses. The Journal of Rheumatology 2014;41(2):194-205.	Guidance on using existing tools
Golder S, Loke YK, Bland M. Unpublished data can be of value in systematic reviews of adverse effects: methodological overview. Journal of Clinical Epidemiology 2010;63(10):1071-81.	Paper does not report on a structured tool
Golder S, Loke YK. Is there evidence for biased reporting of published adverse effects data in pharmaceutical industry-funded studies? British Journal of Clinical Pharmacology 2008;66(6):767-73.	Paper does not report on a structured tool
Goodyear-Smith FA, van Driel ML, Arroll B, Del Mar C. Analysis of decisions made in meta-analyses of depression screening and the risk of confirmation bias: a case study. BMC Med Res Methodol 2012;12:76.	Paper does not report on a structured tool
Grant S, Pedersen ER, Osilla KC, Kulesza M, D'Amico EJ. It is time to develop appropriate tools for assessing minimal clinically important differences, performance bias and quality of evidence in reviews of behavioral interventions. Addiction 2016;111(9):1533-5.	Paper does not report on a structured tool
Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostatistics (Oxford, England) 2001;2(4):463-71.	Describes statistical methods only
Haddaway NR, Woodcock P, Macura B, Collins A. Making literature reviews more reliable through application of lessons from systematic reviews. Conservation Biology 2015;29(6):1596-605.	Guidance on using existing tools
Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. Statistics in Medicine 2000;19(24):3325-36.	Describes statistical methods only
Heck NC, Mirabito LA, LeMaire K, Livingston NA, Flentje A. Omitted data in randomized controlled trials for anxiety and depression: A systematic review of the inclusion of sexual orientation and gender identity. Journal of Consulting and Clinical Psychology 2017;85(1):72-6.	Paper does not report on a structured tool
Higgins JPT, Lane PW, Anagnostelis B, Anzures-Cabrera J, Baker NF, Cappelleri JC, et al. A tool to assess the quality of a meta-analysis. Research Synthesis Methods 2013;4(4):351-66.	Refers to a tool to assess quality of published systematic reviews
Hoy D, Brooks P, Woolf A, Blyth F, March L, Bain C, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. J Clin Epidemiol 2012;65(9):934-9.	Tool does not assess reporting bias
Hsu W, Speier W, Taira RK. Automated extraction of reported statistical analyses: towards a logical representation of clinical trial literature. AMIA Annual Symposium proceedings AMIA Symposium 2012;2012:350-9.	Paper does not report on a structured tool
Ioannidis JPA, Munafo MR, Fusar-Poli P, Nosek BA, David SP. Publication and other reporting biases in cognitive sciences:	Paper does not report on a

Reference	Reason for exclusion
detection, prevalence, and prevention. Trends in Cognitive Sciences 2014;18(5):235-41.	structured tool
Ioannidis JPA, Trikalinos TA. An exploratory test for an excess of significant findings. Clinical Trials 2007;4(3):245-53.	Describes statistical methods only
Ioannidis JPA, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. CMAJ 2007;176(8):1091-6.	Describes statistical methods only
Jarde A, Losilla J-M, Vives J, Rodrigo MF. Q-Coh: A tool to screen the methodological quality of cohort studies in systematic reviews and meta-analysis. International Journal of Clinical and Health Psychology 2013;13(2):138-46.	Tool does not assess reporting bias
Jefferson T, Jones MA, Doshi P, Del Mar CB, Hama R, Thompson MJ, et al. Risk of bias in industry-funded oseltamivir trials: comparison of core reports versus full clinical study reports. BMJ Open 2014;4(9):e005253.	No psychometric properties assessed
Johnson BT, Low RE, MacDonald HV. Panning for the gold in health research: incorporating studies' methodological quality in meta-analysis. Psychology & Health 2015;30(1):135-52.	Describes statistical methods only
Johnston BC, Patrick DL, Busse JW, Schunemann HJ, Agarwal A, Guyatt GH. Patient-reported outcomes in meta-analyses--Part 1: assessing risk of bias and combining outcomes. Health and Quality of Life Outcomes 2013;11:109.	Guidance on using existing tools
Jorgensen L, Paludan-Muller AS, Laursen DR, Savovic J, Boutron I, Sterne JA, et al. Evaluation of the Cochrane tool for assessing risk of bias in randomized clinical trials: overview of published comments and analysis of user practice in Cochrane and non-Cochrane reviews. Syst Rev 2016;5:80.	No psychometric properties assessed
Jurgens T, Whelan AM, MacDonald M, Lord L. Development and evaluation of an instrument for the critical appraisal of randomized controlled trials of natural products. BMC Complement Altern Med 2009;9:11.	Tool does not assess reporting bias
Jurgens TM, Whelan AM. Development and evaluation of an instrument for the critical appraisal of randomized controlled trials of natural products. Canadian Journal of Hospital Pharmacy 2011;64(1):68.	Tool does not assess reporting bias
Katikireddi SV, Egan M, Petticrew M. How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. Journal of Epidemiology and Community Health 2015;69(2):189-95.	Audit of tools used in systematic reviews
Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar S, Grimmer KA. A systematic review of the content of critical appraisal tools. BMC Med Res Methodol 2004;4:22.	Systematic review of tools
Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in	Describes statistical methods only

Reference	Reason for exclusion
systematic reviews. Statistics in Medicine 2012;31(20):2179-95.	
Kocsis JH, Gerber AJ, Milrod B, Roose SP, Barber J, Thase ME, et al. A new scale for assessing the quality of randomized clinical trials of psychotherapy. Comprehensive Psychiatry 2010;51(3):319-24.	Tool does not assess reporting bias
Kovacs FM, Abraira V. Language Bias in a Systematic Review of Chronic Pain: How to Prevent the Omission of Non-English Publications? The Clinical Journal of Pain 2004;20(3):199-200.	Paper does not report on a structured tool
Krauth D, Woodruff TJ, Bero L. Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. Environmental Health Perspectives 2013;121(9):985-92.	Systematic review of tools
Kromrey JD, Rendina-Gobioff G. On Knowing What We Do Not Know: An Empirical Comparison of Methods to Detect Publication Bias in Meta-Analysis. Educational and Psychological Measurement 2006;66(3):357-73.	Describes statistical methods only
Lamont RF. A quality assessment tool to evaluate tocolytic studies. BJOG 2006;113(Suppl 3):96-9.	Tool does not assess reporting bias
Langendam M, Carrasco-Labra A, Santesso N, Mustafa RA, Brignardello-Petersen R, Ventresca M, et al. Improving GRADE evidence tables part 2: A systematic survey of explanatory notes shows more guidance is needed. J Clin Epidemiol 2016;74:19-27.	No psychometric properties assessed
Liebherz S, Schmidt N, Rabung S. How to assess the quality of psychotherapy outcome studies: A systematic review of quality assessment criteria. Psychotherapy Research 2016;26(5):573-89.	Systematic review of tools
Liebherz S, Schmidt N, Rabung S. Study Quality and its Influence on Treatment Outcome in Studies on the Effectiveness of Inpatient Psychotherapy - A Meta-Analysis. PPmP Psychotherapie Psychosomatik Medizinische Psychologie 2016;66(1):31-8.	Not written in English
Lohr KN, Carey TS. Assessing "best evidence": issues in grading the quality of studies for systematic reviews. The Joint Commission Journal on Quality Improvement 1999;25(9):470-9.	Guidance on using existing tools
Lonjon G, Porcher R, Ergina P, Fouet M, Boutron I. Potential Pitfalls of Reporting and Bias in Observational Studies With Propensity Score Analysis Assessing a Surgical Procedure: A Methodological Systematic Review. Ann Surg 2016:no pagination.	Paper does not report on a structured tool
Lundh A, Gotzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. BMC Med Res Methodol 2008;8:22.	Guidance on using existing tools
Lynch HN, Goodman JE, Tabony JA, Rhomberg LR. Systematic comparison of study quality criteria. Regul Toxicol Pharmacol 2016;76:187-98.	Systematic review of tools
Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, et al. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. PLoS Biology 2015;13(10):e1002273.	Tool does not assess reporting bias

Reference	Reason for exclusion
Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003;83(8):713-21.	Tool does not assess reporting bias
Malmivaara A. Methodological considerations of the GRADE method. Annals of Medicine 2015;47(1):1-5.	Guidance on using existing tools
Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. Journal of the American Medical Informatics Association 2016;23(1):193-201.	Model to semi-automate Cochrane risk of bias tool
McDonagh MS, Peterson K, Balshem H, Helfand M. US Food and Drug Administration documents can provide unpublished evidence relevant to systematic reviews. Journal of Clinical Epidemiology 2013;66(10):1071-81.	Paper does not report on a structured tool
McShane BB, Bockenholt U, Hansen KT. Adjusting for Publication Bias in Meta-Analysis: An Evaluation of Selection Methods and Some Cautionary Notes. Perspectives on Psychological Science 2016;11(5):730-49.	Describes statistical methods only
Millard LAC, Flach PA, Higgins JPT. Machine learning to assist risk-of-bias assessments in systematic reviews. International Journal of Epidemiology 2016;45(1):266-77.	Model to semi-automate Cochrane risk of bias tool
Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Controlled Clinical Trials 1995;16(1):62-73.	Systematic review of tools
Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLoS Med 2014;11(10):e1001744.	Refers to a tool to assess quality of published systematic reviews
Moyer A, Finney JW. Rating methodological quality: toward improved assessment and investigation. Accountability in Research 2005;12(4):299-313.	Guidance on using existing tools
Mueller KF, Briel M, Strech D, Meerpohl JJ, Lang B, Motschall E, et al. Dissemination bias in systematic reviews of animal research: a systematic review. PloS One 2014;9(12):e116016.	Paper does not report on a structured tool
Mueller KF, Meerpohl JJ, Briel M, Antes G, von Elm E, Lang B, et al. Detecting, quantifying and adjusting for publication bias in meta-analyses: protocol of a systematic review on methods. Systematic Reviews 2013;2:60.	Describes statistical methods only
Mueller KF, Meerpohl JJ, Briel M, Antes G, von Elm E, Lang B, et al. Methods for detecting, quantifying, and adjusting for dissemination bias in meta-analysis are described. J Clin Epidemiol 2016;80:25-33.	Describes statistical methods only
Nakagawa S, Noble DWA, Senior AM, Lagisz M. Meta-evaluation of meta-analysis: ten appraisal questions for biologists. BMC Biology 2017;15(1):18.	Refers to a tool to assess quality of published systematic reviews
Nolting A, Perleth M, Langer G, Meerpohl JJ, Gartlehner G, Kaminski-	Not written in English

Reference	Reason for exclusion
Hartenthaler A, et al. [GRADE guidelines: 5. Rating the quality of evidence: publication bias]. Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen 2012;106(9):670-6.	
Norris SL, Moher D, Reeves BC, Shea B, Loke Y, Garner S, et al. Issues relating to selective reporting when including non-randomized studies in systematic reviews on the effects of healthcare interventions. Res Synth Methods 2013;4(1):36-47.	Guidance on using existing tools
Nurmatov UB, Xiong T, Kroes MA. Evaluation of quality assessment tools for non-randomised controlled trials assessing surgical interventions: A systematic review of systematic reviews. Value in Health 2015;18(7):A722.	Systematic review of tools
Odierna DH, Forsyth SR, White J, Bero LA. The cycle of bias in health research: a framework and toolbox for critical appraisal training. Accountability in Research 2013;20(2):127-41.	Paper does not report on a structured tool
Palma Perez S, Delgado Rodriguez M. [Practical considerations on detection of publication bias]. Gac Sanit 2006;20(Suppl 3):10-6.	Not written in English
Pearson M, Peters J. Outcome reporting bias in evaluations of public health interventions: evidence of impact and the potential role of a study register. Journal of Epidemiology and Community Health 2012;66(4):286-9.	No psychometric properties assessed
Petticrew M, Egan M, Thomson H, Hamilton V, Kunkler R, Roberts H. Publication bias in qualitative research: what becomes of qualitative research presented at conferences? Journal of Epidemiology and Community Health 2008;62(6):552-4.	Paper does not report on a structured tool
Pigott TD, Valentine JC, Polanin JR, Williams RT, Canada DD. Outcome-Reporting Bias in Education Research. Educational Researcher 2013;42(8):424-32.	Paper does not report on a structured tool
Pirracchio R, Resche-Rigon M, Chevret S, Journois D. Do simple screening statistical tools help to detect reporting bias? Annals of Intensive Care 2013;3(1):29.	Describes statistical methods only
Quigley JM, Thompson J, Halfpenny N, Scott DA. Critical appraisal of non-randomized controlled trials-a review of recommended and commonly used tools. Value in Health 2014;17(3):A203.	Systematic review of tools
Quigley JM, Thompson JC, Halfpenny NJ, Scott DA. Critical appraisal of real world evidence-a review of recommended and commonly used tools. Value in Health 2015;18(7):A684.	Systematic review of tools
Quintana DS. From pre-registration to publication: A non-technical primer for conducting a meta-analysis to synthesize correlational data. Front Psychol 2015;6:1549.	Paper does not report on a structured tool
Rangel SJ, Kelsey J, Colby CE, Anderson J, Moss RL. Development of a quality assessment scale for retrospective clinical studies in pediatric surgery. Journal of Pediatric Surgery 2003;38(3):390-6.	Tool does not assess reporting bias
Rosella L, Bowman C, Pach B, Morgan S, Fitzpatrick T, Goel V. The development and validation of a meta-tool for quality appraisal of	Tool does not assess reporting bias

Reference	Reason for exclusion
public health evidence: Meta Quality Appraisal Tool (MetaQAT). Public Health 2016 Jul;136:57-65.	
Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. International Journal of Epidemiology 2007;36(3):666-76.	Systematic review of tools
Santaguida PL, Riley CM, Matchar DB. Chapter 5: Assessing risk of bias as a domain of quality in medical test studies. Journal of General Internal Medicine 2012;27(Suppl 1):S33-S8.	Guidance on using existing tools
Savovic J, Weeks L, Sterne JA, Turner L, Altman DG, Moher D, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. Syst Rev 2014;3:37.	No psychometric properties assessed
Seehra J, Pandis N, Koletsi D, Fleming PS. Use of quality assessment tools in systematic reviews was varied and inconsistent. J Clin Epidemiol 2016;69:179-84.e5.	Audit of tools used in systematic reviews
Shamliyan T, Kane RL, Dickinson S. A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. Journal of Clinical Epidemiology 2010;63(10):1061-70.	Systematic review of tools
Shamliyan TA, Kane RL, Ansari MT, Raman G, Berkman ND, Grant M, et al. Development quality criteria to evaluate nontherapeutic studies of incidence, prevalence, or risk factors of chronic diseases: pilot study of new checklists. Journal of Clinical Epidemiology 2011;64(6):637-57.	Tool does not assess reporting bias
Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10.	Refers to a tool to assess quality of published systematic reviews
Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. Journal of Clinical Epidemiology 2009;62(10):1013-20.	Refers to a tool to assess quality of published systematic reviews
Shuang M, Zhao C, Zhang L, Shang HC. Using SYRCLE tools to evaluate the methodological quality of animal experiments of stroke in China. Chinese Journal of Evidence-Based Medicine 2016;16(5):592-7.	Not written in English
Singh S, Khosla S. Suboptimal choice of methodology for meta-analysis and publication bias assessment. The American Journal of Cardiology 2015;115(12):1782-3.	Describes statistical methods only
Smyth RM, Kirkham JJ, Jacoby A, Altman DG, Gamble C, Williamson PR. Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists. BMJ 2011;342:c7153.	Paper does not report on a structured tool
Sohani ZN, Meyre D, de Souza RJ, Joseph PG, Gandhi M, Dennis BB, et al. Assessing the quality of published genetic association studies in	Tool does not assess

Reference	Reason for exclusion
meta-analyses: the quality of genetic studies (Q-Genie) tool. BMC Genet 2015;16:50.	reporting bias
Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technology Assessment (Winchester, England) 2010;14(8):iii-193.	Paper does not report on a structured tool
Spooner CH, Pickard AS, Menon D. Edmonton Quality Assessment Tool for Drug Utilization Reviews: EQUATDUR-2: the development of a scale to assess the methodological quality of a drug utilization review. Medical Care 2000;38(9):948-58.	Tool does not assess reporting bias
Tate RL, Perdices M, Rosenkoetter U, Wakim D, Godbee K, Togher L, et al. Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: the 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. Neuropsychological Rehabilitation 2013;23(5):619-38.	Tool does not assess reporting bias
Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters M, et al. AHRQ Methods for Effective Health Care Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012.	Guidance on using existing tools
Voss PH, Rehfuess EA. Quality appraisal in systematic reviews of public health interventions: an empirical study on the impact of choice of tool on meta-analysis. Journal of Epidemiology and Community Health 2013;67(1):98-104.	Evaluation of existing tools
Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2008. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed 7/03/2017).	Tool does not assess reporting bias
Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J. A systematic review finds that diagnostic reviews fail to incorporate quality despite available tools. Journal of Clinical Epidemiology 2005;58(1):1-12.	Systematic review of tools
Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.	Tool does not assess reporting bias
Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine 2011;155(8):529-36.	Tool does not assess reporting bias
Wiart L, Kolaski K, Vogtle LK, Butler C, Romeiser Logan L, Hickman R, et al. Inter-rater reliability and concurrent validity of the AACPDM study design and quality rating system for conducting systematic	Refers to a tool to assess quality of published systematic reviews

Reference	Reason for exclusion
reviews (group design). Dev Med Child Neurol 2011;53:74.	

Table S3. General characteristics of included tools

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
Balshem 2013[1]	AHRQ outcome and analysis reporting bias framework	Domain-based	Reporting bias only	Bias due to selective non-reporting and bias in selection of the reported result	Randomized trials	Specific outcome/ result in a study	Expert consensus (via email)	Brief annotation per item/response option	No
Berkman 2013[2]	AHRQ tool for evaluating the risk of reporting bias	Domain-based	Reporting bias only	Bias due to selective publication and bias due to selective non-reporting	Systematic reviews	Specific synthesis of studies	Not stated	Brief annotation per item/response option	No
Downes 2016[3]	AXIS tool (Appraisal tool for Cross-Sectional Studies)	Checklist	Multiple sources of bias	Bias due to selective non-reporting	Cross-sectional studies	Whole study	Literature review, piloting, Delphi study	None	No
Downs 1998[4]	Downs-Black tool	Scale	Multiple sources of bias	Bias in selection of the	Randomized trials and non-	Whole study	Literature review, piloting,	Brief annotation per item/response	Yes

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
				reported result	randomized studies of interventions		psychometric testing	option	
Guyatt 2011[5-9]	GRADE	Domain-based	Multiple sources of bias	Bias due to selective publication and bias due to selective non-reporting	Systematic reviews	Specific synthesis of studies	Literature review, expert consensus (face-to-face and email), user testing	Detailed guidance manual	Yes
Hayden 2013[10]	QUIPS (Quality In Prognosis Studies) tool	Domain-based	Multiple sources of bias	Bias due to selective non-reporting	Prognosis studies	Whole study	Modified Delphi approach, nominal group technique at facilitated discussion workshop; piloting	Brief annotation per item/response option	Yes
Higgins 2011[11-13]	Cochrane risk of bias tool for randomized trials	Domain-based	Multiple sources of bias	Bias due to selective non-reporting and bias in selection of the reported	Randomized trials	Whole study	Literature review, informal consensus at facilitated meeting, piloting, focus groups and	Detailed guidance manual	Yes

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
				result			surveys, followed by consensus meeting		
Higgins 2016[14 15]	RoB 2.0 (revised tool for assessing risk of bias in randomized trials)	Domain-based	Multiple sources of bias	Bias in selection of the reported result	Randomized trials	Specific outcome/ result in a study	Literature review, informal consensus at facilitated meeting, piloting	Detailed guidance manual	No
Hoojimans 2014[16]	SYRCLE's RoB tool (SYstematic Review Centre for Laboratory animal Experimentation)	Domain-based	Multiple sources of bias	Bias due to selective non-reporting and bias in selection of the reported result	Animal studies	Whole study	Adaptation of existing tool, literature review	Brief annotation per item/response option	No
Kim 2013[17]	RoBANS (Risk of Bias Assessment Tool for Nonrandomized Studies)	Domain-based	Multiple sources of bias	Bias due to selective non-reporting and bias in selection of the	Non-randomized studies of interventions	Whole study	Literature review, psychometric testing	Brief annotation per item/response option	Yes

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
				reported result					
Kirkham 2010[18 19]	ORBIT-I (Outcome Reporting Bias In Trials) classification system for benefit outcomes	Domain-based	Reporting bias only	Bias due to selective non-reporting	Randomized trials	Specific outcome/ result in a study	Iteratively developed as part of a methodological study	Worked example for each response option	Yes
Meader 2014[20 21]	SAQAT (Semi-Automated Quality Assessment Tool)	Domain-based	Multiple sources of bias	Bias due to selective publication and bias due to selective non-reporting	Systematic reviews	Specific synthesis of studies	Development of logic model based on GRADE articles and piloting	None	Yes
Reid 2015[22]	Selective reporting bias algorithm	Domain-based	Reporting bias only	Bias due to selective non-reporting and bias in selection of the reported	Randomized trials	Whole study	Not stated	Brief annotation per item/response option	No

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
				result					
Saini 2014[23]	ORBIT-II (Outcome Reporting Bias In Trials) classification system for harm outcomes	Domain-based	Reporting bias only	Bias due to selective non-reporting	Randomized trials and non-randomized studies of interventions	Specific outcome/ result in a study	Iteratively developed as part of a methodological study	Worked example for each response option	No
Salanti 2014[24 25]	Framework for evaluating the quality of evidence from a network meta-analysis	Domain-based	Multiple sources of bias	Bias due to selective publication and bias due to selective non-reporting	Network meta-analyses	Specific synthesis of studies	Adaptation of existing tool	Detailed annotation per item/response option	No
Sterne 2016[26]	ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool	Domain-based	Multiple sources of bias	Bias in selection of the reported result	Non-randomized studies of interventions	Specific outcome/ result in a study	Expert consensus meetings (face-to-face), piloting	Detailed guidance manual	Yes

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
Viswanathan 2012[27]	RTI Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures	Domain-based	Multiple sources of bias	Bias due to selective non-reporting	Non-randomized studies of interventions or exposures	Whole study	Literature review, expert consensus (via email), cognitive testing, psychometric testing	Brief annotation per item/response option	No
Viswanathan 2013[28]	RTI Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures	Domain-based	Multiple sources of bias	Bias due to selective non-reporting	Non-randomized studies of interventions or exposures	Whole study	Literature review, expert consensus (via email)	Brief annotation per item/response option	No

References

1. Balshem H, Stevens A, Ansari M, et al. Finding grey literature evidence and assessing for outcome and analysis reporting biases when comparing medical interventions: AHRQ and the Effective Health Care Program. (Prepared by the Oregon Health and Science University and the University of Ottawa Evidence-based Practice Centers under Contract Nos. 290-2007-10057-I and 290-2007-10059-I.) AHRQ Publication No. 13(14)-EHC096-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

2. Berkman ND, Lohr KN, Ansari M, et al. Chapter 15 Appendix A: A Tool for Evaluating the Risk of Reporting Bias (in Chapter 15: Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update). Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I). AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm

3. Downes MJ, Brennan ML, Williams HC, et al. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ open* 2016;6:e011458.

4. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.

5. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.

6. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64(12):1277-82.

7. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol* 2011;64(4):407-15.

8. Schünemann H, Brożek J, Guyatt G, et al. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. [Updated October 2013]. Available from http://gdt.guidelinedevelopment.org/app/handbook/handbook.html.

9. Santesso N, Carrasco-Labra A, Langendam M, et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. *J Clin Epidemiol* 2016

10. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.

11. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons 2008:187-241.

12. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org/.

13. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

14. Higgins JPT, Savović J, Page MJ, et al. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0), Version 20 October 2016. Available from http://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/rob2-0/ [accessed 19 September 2017].

15. Higgins JPT, Sterne JAC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials. *Cochrane Methods Cochrane Database of Systematic Reviews* 2016;10(Suppl 1):29-31.

16. Hooijmans CR, Rovers MM, de Vries RBM, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.

17. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013;66(4):408-14.

18. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.

19. Dwan K, Gamble C, Kolamunnage-Dona R, et al. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials* 2010;11:52.

20. Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic reviews* 2014;3(1):82.

21. Stewart GB, Higgins JP, Schunemann H, et al. The use of Bayesian networks to assess the quality of evidence from research synthesis: 1. *PLoS One* 2015;10(3):e0114497.

22. Reid EK, Tejani AM, Huan LN, et al. Managing the incidence of selective reporting bias: a survey of Cochrane review groups. *Systematic reviews* 2015;4:85.

23. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ* 2014;349:g6501.

24. Salanti G, Giovane CD, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9(7):e99682.

25. Higgins JP, Del Giovane C, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *Value Health* 2014;17(7):A324.

26. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.

27. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* 2012;65(2):163-78.

28. Viswanathan M, Berkman ND, Dryden DM, et al. AHRQ Methods for Effective Health Care. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or

Exposures: Further Development of the RTI Item Bank. Rockville (MD): Agency for Healthcare Research and Quality (US) 2013.

Table S4. Items and response options relating to risk of reporting biases

Article ID	Tool	Items	Response options
Balshem 2013[1]	AHRQ outcome and analysis reporting bias framework	1. Across all study source documents, what is the risk of ORB/ARB? Compare published report(s) against (1) study protocol (if not retrieved in literature search), (2) trial registry entry/regulatory documents/industry documents, (3) other sources if applicable. 2. If ORB risk unclear: Given the study objectives, duration, and other investigated outcomes, could the study have also likely measured the outcome of interest but not reported it?	**Outcome reporting bias risk positive (ORB risk +)**: If reviewers determine that an outcome X was planned but the results were not reported, or were only partially reported in study documents, then the study is at risk of reporting bias for that outcome ("ORB risk +"). Also, if reviewers determine that an outcome X was not planned but the results were reported, then the study is at risk of reporting bias for that outcome ("ORB risk +"). Also, for studies for which the risk of reporting bias cannot be ruled out, reviewers should ask the question: "Given the study objectives, duration, and other investigated outcomes, could the study have also likely measured the outcome of interest but not reported it?" When the answer is "yes" (e.g., another reported outcome in the study leads the reviewer to believe that outcome X would have been collected), then the study should be rated "ORB risk +" for that outcome. **Outcome reporting bias risk negative (ORB risk -)**: When it is clear to the reviewers that outcome X was planned (e.g. from protocol, regulatory submissions, etc.), complete outcome data are available from at least one study document (published or otherwise), and the outcome was appropriately analyzed as planned, then the study is not at risk for reporting bias for this outcome. Also, for studies for which the risk of reporting bias cannot be ruled out, reviewers should ask the question: "Given the study objectives, duration, and other investigated outcomes, could the study have also

Article ID	Tool	Items	Response options
			likely measured the outcome of interest but not reported it?" If the answer is "no" the study should be rated as "ORB risk–". **Outcome reporting bias risk unclear (ORB risk unclear)**: If the reviewers are unable to determine whether an outcome X was planned, but data are reported completely or partially, then the study risk of outcome and analysis reporting bias may be categorized as "unclear". This would also apply to a study that did not report any outcome of review interest across all source documents but was eligible on population, intervention, comparator, and other criteria. Also, for studies for which the risk of reporting bias cannot be ruled out, reviewers should ask the question: "Given the study objectives, duration, and other investigated outcomes, could the study have also likely measured the outcome of interest but not reported it?" If it still remains unclear whether the outcome of interest may have been assessed, the study should be categorized as "ORB risk unclear." **Analysis reporting bias risk positive (ARB risk +)**: When reported results are based on a different analysis, effect measure, cut-off, etc. than what was prespecified, then the study is at risk of analysis reporting bias for that outcome ("ARB risk +"). A study is also at risk of analysis reporting ("ARB risk +") because there is no way to know whether the reported analysis was planned or post hoc. **Analysis reporting bias risk negative (ARB risk -)**: When it is clear to the reviewers that outcome X was planned (e.g. from protocol, regulatory submissions, etc.),

Article ID	Tool	Items	Response options
			complete outcome data are available from at least one study document (published or otherwise), and the outcome was appropriately analyzed as planned, then the study is not at risk for reporting bias for this outcome **Analysis reporting bias risk unclear (ARB risk unclear)**: If the reviewers are unable to determine whether an outcome X was planned, but data are reported completely or partially, then the study risk of outcome and analysis reporting bias may be categorized as “unclear”. This would also apply to a study that did not report any outcome of review interest across all source documents but was eligible on population, intervention, comparator, and other criteria.
Berkman 2013[2]	AHRQ tool for evaluating the risk of reporting bias	1. Are all the following criteria met: ≥10 studies contributing data for an outcome, studies of unequal sizes, no substantial clinical and methodological differences between smaller and larger studies, and quantitative results accompanied with measures of dispersion? 2. If yes, do smaller studies tend to demonstrate more favorable results? (visual assessment) 3. If yes, what is the result of a test for funnel plot asymmetry? 4. If test is positive, would a clinical decision differ for estimates from a fixed effects versus random effect model because the findings from a fixed effect model are closer to the null? 5. If no to the first question, is there an explanation for substantial heterogeneity? 6. If no to any of Q1-5, what is the estimated N of studies that are affected by SOR, SAR,	**Suspected risk of reporting bias**: Testing for funnel plot asymmetry demonstrates a substantial likelihood of bias, and/or a qualitative assessment suggests the likelihood of missing studies, analyses, or outcomes data that may alter the conclusions from the reported evidence. **Undetected risk of reporting bias**: All alternative scenarios.

Article ID	Tool	Items	Response options
		nonpublication, or nonaccessibility? 7. If no to any of Q1-5, what is the total sample size of evidence affected by reporting bias (when known)? 8. If no to any of Q1-5, what is the total N of studies in evidence base? 9. If no to any of Q1-5, what is the total N of participants in evidence base? 10. If no to any of Q1-5, what is the consistency of effect estimates across contributing studies? 11. If no to any of Q1-5, what are the study limitations for the evidence base? 12. If no to any of Q1-5, what is the comprehensiveness of study retrieval and identification?	
Downes 2016[3]	AXIS tool (Appraisal tool for Cross-Sectional Studies)	1. Were the results for the analyses described in the methods, presented?	**Yes**: Not stated **No**: Not stated **Do not know/comment**: Not stated
Downs 1998[4]	Downs-Black tool	1. If any of the results of the study were based on "data dredging", was this made clear?	**Yes**: Any analyses that had not been planned at the outset of the study were clearly indicated. Also, no retrospective unplanned subgroup analyses were reported. **No**: Any analyses that had not been planned at the outset of the study were not clearly indicated. **Unable to determine**: Not stated
Guyatt 2011[5-9]	GRADE	1. Study limitations (including selective outcome reporting) 2. Publication bias	**Study limitations domain – No serious limitations, do not downgrade**: Most information is from studies at low risk of bias (i.e. those with low risk of bias for all key criteria, including lack of allocation concealment, lack of blinding, incomplete accounting of patients and outcome

Article ID	Tool	Items	Response options
			events, selective outcome reporting bias [incomplete or absent reporting of some outcomes and not others on the basis of the results], other limitations [stopping early for benefit, use of unvalidated outcome measures, carryover effects in crossover trial, recruitment bias in cluster-randomized trial]) **Study limitations domain – Serious limitations, rate down one level (i.e., from high to moderate quality)**: Most information is from studies at moderate risk of bias **Study limitations domain – Very serious limitations, rate down two levels (i.e., from high to low quality or moderate to very low)**: Most information is from studies at high risk of bias **Publication bias domain – Undetected**: None of the criteria for "strongly suspected" are met **Publication bias domain – Strongly suspected**: "In general, review authors and guideline developers should consider rating down for likelihood of publication bias when the evidence consists of a number of small studies. The inclination to rate down for publication bias should increase if most of those small studies are industry sponsored or likely to be industry sponsored (or if the investigators share another conflict of interest)...Another criterion for publication bias is the pattern of study results. Suspicion may increase if visual inspection demonstrates an asymmetrical rather than a symmetrical funnel plot or if statistical tests of asymmetry are positive. Although funnel plots may be helpful, review authors and guideline developers should bear in mind that visual assessment of funnel plots is

Article ID	Tool	Items	Response options
			distressingly prone to error. Enhancements of funnel plots may (or may not) help to improve reproducibility and validity associated with their use...Furthermore, systematic review and guideline authors should bear in mind that even if they find convincing evidence of asymmetry, publication bias is not the only explanation. For instance, if smaller studies suffer from greater study limitations, they may yield biased overestimates of effects. Another explanation would be that, because of a more restrictive (and thus responsive) population, or a more careful administration of the intervention, the effect may actually be larger in the small studies...More compelling than any of these theoretical exercises is authors' success in obtaining the results of some unpublished studies and demonstrating that the published and unpublished data show different results. In these circumstances, the possibility of publication bias looms large. The risk of publication bias is probably larger for observational studies than for RCTs, particularly small observational studies and studies conducted on data collected automatically (e.g. in the electronic medical record or in a diabetes registry) or data collected for a previous study. In these instances, it is difficult for the reviewer to know if the observational studies that appear in the literature represent all or a fraction of the studies conducted, and whether the analyses in them represent all or a fraction of those conducted. In these instances, reviewers may consider the risk of publication bias as substantial" [6]. "Guideline panels and authors of systematic reviews should consider the extent to which they are uncertain about the magnitude of the effect due to selective publication

Article ID	Tool	Items	Response options
			of studies and they may downgrade the quality of evidence by one level. Consider: study design (experimental vs. observational); study size (small studies vs. large studies); lag bias (early publication of positive results); search strategy (was it comprehensive?); asymmetry in funnel plot" [8]. "Relevant content: whether publication bias is undetected or suspected; interpretation of funnel plot; comprehensiveness of the search strategies and methods to identify all available evidence; presence of small (often positive) studies with for profit interest...Indicate the reason publication bias is detected (e.g. asymmetrical funnel plot, small studies with positive results, suspected selective availability of data from published, or unpublished studies)" [9].
Hayden 2013[10]	QUIPS (Quality In Prognosis Studies) tool	1. Statistical analysis and reporting (the statistical analysis is appropriate and all primary outcomes are reported). Prompting items include (a) Sufficient presentation of data to assess the adequecy of the analytic strategy; (b) Strategy for model building is appropriate and is based on a conceptual framework or model; (c) The selected statistical model is adequate for the design of the study; (d) There is no selective reporting of results.	**Low risk of bias**: The reported results are unlikely to be spurious or biased related to analysis or reporting **Moderate risk of bias**: The reported results may be spurious or biased related to analysis or reporting **High risk of bias**: The reported results are very likely to be spurious or biased related to analysis or reporting
Higgins 2011[11-13]	Cochrane risk of bias tool for randomized trials	1. Are reports of the study free of suggestion of selective outcome reporting? (2008 version); Reporting bias due to selective outcome reporting (2011 version)	**Low risk of bias**: Any of the following – The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way; The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text

Article ID	Tool	Items	Response options
			of this nature may be uncommon). **High risk of bias**: Any one of the following – Not all of the study's pre-specified primary outcomes have been reported; One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not prespecified; One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; The study report fails to include results for a key outcome that would be expected to have been reported for such a study. **Unclear risk of bias**: Insufficient information to permit judgement of 'Low risk' or 'High risk'. It is likely that the majority of studies will fall into this category.
Higgins 2016[14 15]	RoB 2.0	1. Are the reported outcome data likely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain, or from multiple analyses of the data?	**Low risk of bias**: Reported outcome data are unlikely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain, and reported outcome data are unlikely to have been selected, on the basis of the results, from multiple analyses of the data. **High risk of bias**: Reported outcome data are likely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain, or from multiple analyses of the data (or both). **Some concerns**: There is insufficient information

Article ID	Tool	Items	Response options
			available to exclude the possibility that reported outcome data were selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain, or from multiple analyses of the data.
Hoojimans 2014[16]	SYRCLE's RoB tool (SYstematic Review Centre for Laboratory animal Experimentation)	1. Are reports of the study free of selective outcome reporting? Includes two signalling questions: Was the study protocol available and were all of the study's pre-specified primary and secondary outcomes reported in the current manuscript?; Was the study protocol not available, but was it clear that the published report included all expected outcomes (i.e. comparing methods and results section)?	**Low risk of bias**: Not stated, but assume same criteria as Cochrane risk of bias tool for randomized trials [13]. **High risk of bias**: Not all of the study's pre-specified primary outcomes have been reported; One or more primary outcomes have been reported using measurements, analysis methods or data subsets (e.g. subscales) that were not pre-specified in the protocol; One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting has been provided, such as an unexpected adverse effect); The study report fails to include results for a key outcome that would be expected to have been reported for such a study. **Unclear risk of bias**: Not stated, but assume same criteria as Cochrane risk of bias tool for randomized trials [13].
Kim 2013[17]	RoBANS (Risk of Bias Assessment Tool for Nonrandomized Studies)	1. Reporting biases caused by the selective reporting of outcomes	**Low risk of bias**: Any one of the following conditions – The experimental protocol is available, and the pre-defined primary/secondary outcomes were described as planned; All of the expected outcomes were included in the study descriptions (even in the absence of the experimental protocols). **High risk of bias**: Any one of the following conditions – The pre-defined primary outcomes were not fully

Article ID	Tool	Items	Response options
			reported; The outcomes were not reported in accordance with the previously defined standards; Primary outcomes that were not pre-specified in the study existed (except for outcomes with clear explanations, such as unexpected adverse effects); The existence of incomplete reporting regarding the primary outcome of interest; The absence of reports on important outcomes that would be expected to be reported for studies in related fields. **Unclear risk of bias**: It is uncertain whether the selective outcome reporting resulted in a 'high risk' or a 'low risk' of bias.
Kirkham 2010[18 19]	ORBIT-I (Outcome Reporting Bias In Trials) classification system for benefit outcomes	1. The Outcome Reporting Bias In Trials (ORBIT) study classification system for missing or incomplete outcome reporting in reports of randomised trials	**Low risk of bias**: A “low risk” classification was awarded when it was suspected, but not actually known, that the outcome was either not measured, measured but not analysed, or measured and analysed but either partially reported or not reported for a reason unrelated to the results obtained. Specific examples include: (C) Trial report states that outcome was analysed but insufficient data were presented for the trial to be included in meta-analysis or to be considered to be fully tabulated; (F) Clear that outcome was measured but not necessarily analysed, and judgment says unlikely to have been analysed but not reported because of non-significant results; (H) Not mentioned but clinical judgment says outcome unlikely to have been measured at all. **High risk of bias**: A “high risk” classification was awarded when it was either known or suspected that the results were partially or not reported because the treatment comparison was statistically non-significant (P>0.05).

Article ID	Tool	Items	Response options
			Specific examples include: (A) Trial report states that outcome was analysed but only reports that result was not significant (typically stating P>0.05); (D) Trial report states that outcome was analysed but no results reported; (E) Clear that outcome was measured but not necessarily analysed, and judgment says likely to have been analysed but not reported because of non-significant results; (G) Not mentioned but clinical judgment says outcome likely to have been measured and analysed but not reported on the basis of non-significant results. **No risk of bias**: A "no risk" classification was reserved for cases where it was known that the outcome was not measured, known that it was measured but not analysed, or known that it was measured and analysed but the reason for partial or no reporting was not because the results were statistically non-significant. Specific examples include: (B) Trial report states that outcome was analysed but only reports that result was significant (typically stating P<0.05); (I) Clear that outcome was not measured.
Meader 2014[20 21]	SAQAT (Semi-Automated Quality Assessment Tool)	Study limitations domain 1. Were data reported consistently for the outcome of interest (i.e. no potential selective reporting)? Publication bias domain 1. Did the authors conduct a comprehensive search? 2. Did the authors search for grey literature? 3. Authors did not apply restrictions to study selection on the basis of language?	**Study limitations domain – No serious limitations**: No problem for any source of risk of bias. **Study limitations domain – Serious limitations**: Selection bias results in serious limitations, or very serious limitations if combined with a problem from any alternative source; two problems from other sources (e.g. detection bias, attrition bias) result in serious limitations. **Study limitations domain – Very serious limitations**:

Article ID	Tool	Items	Response options
		4. There was no industry influence on studies included in the review? 5. There was no evidence of funnel plot asymmetry? 6. There was no discrepancy in findings between published and unpublished trials?	Selection bias results in serious limitations, or very serious limitations if combined with a problem from any alternative source; three problems result in very serious limitations **Publication bias domain – Strongly suspected**: High probability of publication bias. Responses to each item are entered into a Bayesian network to ascertain the probabilities of each GRADE domain. Publication bias is determined by a combination of discrepancy between published and unpublished studies (yes/no), amount of statistical information (high/intermediate/low), industry influence (yes/no) and search integrity (high/low), with the former carrying greatest weight. That is, the probability of publication bias is always considered high when there is a discrepancy between published and unpublished studies (regardless of responses to other items). **Publication bias domain – Undetected**: Low probability of publication bias (as determined by the Bayesian network described above.
Reid 2015[22]	Selective reporting bias algorithm	1. Protocol available? 2. Trial registration? 3. Outcomes described? 4. Response from contact with study authors? 5. Outcomes match?	**High risk of bias**: Outcomes are described in the protocol or trial registry or by the review authors when contacted, and they do not match the outcomes reported. **Low risk of bias**: Outcomes are described in the protocol or trial registry or by the review authors when contacted, and they do match the outcomes reported. **Unclear risk of bias**: Outcomes are not described in the protocol or trial registry, or a protocol or trial registry are not available and no response is received from review

Article ID	Tool	Items	Response options
			authors when contacted.
Saini 2014[23]	ORBIT-II (Outcome Reporting Bias In Trials) classification system for harm outcomes	1. ORBIT-II classification system	**Low risk of bias**: Specific examples include: (P3) Explicit specific harm measured and compared across treatment groups, although insufficient reporting for meta-analysis or full tabulation; (T1) Clinical judgement says specific harm likely measured but no events, because specific harm not mentioned but all other specific harms fully reported; (T2) Clinical judgement says specific harm likely measured but no events, because there was no description of specific harms; (U) Specific harm outcome not explicitly mentioned, clinical judgment says unlikely measured (no harms mentioned or reported). **High risk of bias**: In the context of harm outcomes, we awarded classifications for "high risk" outcome reporting bias when the specific harm had been measured but the data were presented or suppressed in a way that would mask the harm profile of particular interventions (including providing detail on the seriousness of the harms)—that is, P1, P2, R, and S classifications. Specific examples include: (P1) States outcome analysed but reported only that P>0.05; (P2) States outcome analysed but reported only that P<0.05; (R1) Clear that outcome was measured but no results reported; (R2) Result reported globally across all groups; (R3) Result reported from some groups only; (S1) Clinical judgment says specific harm outcome likely measured and likely compared across treatment groups, but only pooled adverse events reported (could include specific harm outcome); (S2) Clinical judgment says specific harm outcome likely measured and likely compared across

Article ID	Tool	Items	Response options
			treatment groups, but no harms mentioned or reported. **No risk of bias**: Specific examples include: (Q) Clear that explicit specific harm outcome was measured and clear outcome was not compared; (V) Report clearly specifies that data on specific harm of interest was not measured.
Salanti 2014[24 25]	Framework for evaluating the quality of evidence from a network meta-analysis	1. Study limitations (including selective outcome reporting) evaluated in a specific pairwise effect estimated in network meta-analysis: Determine which direct comparisons contribute to estimation of the NMA treatment effect and integrate risk of bias assessments from these into a single judgment. 2. Publication bias evaluated in a specific pairwise effect estimated in network meta-analysis: Non-statistical consideration of likelihood of non-publication of evidence that would inform the pairwise comparison. Plot pairwise estimates on contour-enhanced funnel plot. 3. Study limitations (including selective outcome reporting) evaluated in treatment ranking estimated in network meta-analysis: Integrate risk of bias assessments from each direct comparison to formulate a single overall confidence rating for treatment rankings. 4. Publication bias evaluated in treatment ranking estimated in network meta-analysis: Non-statistical consideration of likelihood of non-publication for each pairwise comparison. If appropriate, plot NMA estimates on a comparison adjusted funnel plot and assess asymmetry.	**Study limitations domain – No serious limitations, do not downgrade**: Use standard GRADE considerations to inform judgment [7]. **Study limitations domain – Serious limitations, rate down one level (i.e., from high to moderate quality)**: Use standard GRADE considerations to inform judgment [7]. **Study limitations domain – Very serious limitations, rate down two levels (i.e., from high to low quality or moderate to very low)**: Use standard GRADE considerations to inform judgment [7]. **Publication bias domain (evaluated in a specific pairwise effect estimated in network meta-analysis) – Undetected**: Use standard GRADE to inform judgment [6]. **Publication bias domain (evaluated in a specific pairwise effect estimated in network meta-analysis) – Strongly suspected**: "Even after a meticulous search for studies, publication bias can occur and usually it tends to lead to overestimation of an active treatment's effect compared with placebo or other reference treatment. Several approaches have been proposed to generate assumptions about the presence of publication bias, including funnel plots, regression methods and selection models, but each has limitations and their

Article ID	Tool	Items	Response options
			appropriateness is often debated. Making judgements about the presence of publication bias in a network meta-analysis is usually difficult. We suggest that for each observed pairwise comparison, judgements about the presence of publication bias are made using standard GRADE. We recommend that the primary considerations are non-statistical (by considering how likely it is that studies may have been performed but not published) and we advocate the use of contour-enhanced funnel plots, which may help in identifying publication bias as a likely explanation of funnel plot asymmetry. Then, judgements about the direct effects can be summarized to infer about the network estimates by taking into account the contributions of each direct piece of evidence" [24]. **Publication bias domain (evaluated in treatment ranking estimated in network meta-analysis) – Undetected**: Use standard GRADE to inform judgment [6]. **Publication bias domain (evaluated in treatment ranking estimated in network meta-analysis) – Strongly suspected**: "Judgments about the potential impact of publication bias in the ranking of the treatments require, as before, consideration of the comprehensiveness of the search for studies and the likelihood that studies may have been conducted and not published. A statistical approach to detecting bias is offered in certain situations by the comparison-adjusted funnel plot for a network of treatments. In such a plot, the vertical axis represents the inverted standard error of the effect sizes as in a standard funnel plot. However, the horizontal axis represents an adjusted effect size, presenting the

Article ID	Tool	Items	Response options
			difference between each observed effect size and the mean effect size for the specific comparison being made. The use of such a plot is informative only when the comparisons can confidently be ordered in a meaningful way; for example, if all comparisons are of active treatment versus placebo, or all are of a new versus an old drug. Examination of any asymmetry in the plot can help to infer about the possible presence of an association between study size and study effect. Asymmetry does not provide evidence of publication bias, however, since associations between effect size and study size can be due to study limitations or genuine heterogeneity of effects” [24].
Sterne 2016[26]	ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool	1. Is the reported effect estimate likely to be selected, on the basis of the results, from multiple outcome measurements within the outcome domain, multiple analyses of the intervention-outcome relationship, or different subgroups?	**Low risk of bias**: There is clear evidence (usually through examination of a pre-registered protocol or statistical analysis plan) that all reported results correspond to all intended outcomes, analyses and subcohorts. **Moderate risk of bias**: (i) The outcome measurements and analyses are consistent with an a priori plan; or are clearly defined and both internally and externally consistent; and (ii) There is no indication of selection of the reported analysis from among multiple analyses; and (iii) There is no indication of selection of the cohort or subgroups for analysis and reporting on the basis of the results. **Serious risk of bias**: (i) Outcomes are defined in different ways in the methods and results sections, or in different publications of the study; or (ii) There is a high risk of selective reporting from among multiple analyses; or (iii) The cohort or subgroup is selected from a larger study

Article ID	Tool	Items	Response options
			for analysis and appears to be reported on the basis of the results. **Critical risk of bias**: (i) There is evidence or strong suspicion of selective reporting of results; and (ii) The unreported results are likely to be substantially different from the reported results. **No information**: There is too little information to make a judgement (for example, if only an abstract is available for the study).
Viswanathan 2012[27]	RTI Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures	1. Are any important primary outcomes missing from the results? 2. Are any important harms or adverse events that may be a consequence of the intervention/exposure missing from the results?	**Yes (for item on primary outcome)**: No specific criteria stated. Only guidance is "Identify all primary outcomes, including timing of measurement, that one would expect to be reported in the study" **No (for item on primary outcome)**: No specific criteria stated. **Cannot determine (for item on primary outcome)**: No specific criteria stated. **Yes (for item on harm outcome)**: No specific criteria stated. Only guidance is "Identify all important harms, including timing of measurement, that one would expect be reported in the study. Drop if not relevant to body of literature." **Partially (for item on harm outcome)**: No specific criteria stated. **No (for item on harm outcome)**: No specific criteria stated. **Assessment of harms not applicable to this study (for

Article ID	Tool	Items	Response options
			item on harm outcome): No specific criteria stated.
Viswanathan 2013[28]	RTI Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures	1. Are any important primary outcomes missing from the results? 2. Are any important harms or adverse events that may be a consequence of the intervention/exposure missing from the results?	**Yes, important outcome(s) missing (for item on primary outcome)**: No specific criteria stated. Only guidance is "Identify all primary outcomes that one would expect to be reported in the study, including timing of measurement." **No important outcome (s) missing (for item on primary outcome)**: No specific criteria stated. **Cannot determine (for item on primary outcome)**: No specific criteria stated. **Yes, important outcomes missing (for item on harm outcome)**: No specific criteria stated. Only guidance is "Identify all important harms that one would expect be reported in the study, including timing of measurement. Drop if not relevant to body of literature." **No important outcomes missing (for item on harm outcome)**: No specific criteria stated. **Assessment of harms not applicable to this study (for item on harm outcome)**: No specific criteria stated.

References

1. Balshem H, Stevens A, Ansari M, et al. Finding grey literature evidence and assessing for outcome and analysis reporting biases when comparing medical interventions: AHRQ and the Effective Health Care Program. (Prepared by the Oregon Health and Science University and the University of Ottawa Evidence-based Practice Centers under Contract Nos. 290-2007-10057-I and 290-2007-10059-I.) AHRQ Publication No. 13(14)-EHC096-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

2. Berkman ND, Lohr KN, Ansari M, et al. Chapter 15 Appendix A: A Tool for Evaluating the Risk of Reporting Bias (in Chapter 15: Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update). Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I). AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm

3. Downes MJ, Brennan ML, Williams HC, et al. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ open* 2016;6:e011458.

4. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.

5. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.

6. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64(12):1277-82.

7. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol* 2011;64(4):407-15.

8. Schünemann H, Brożek J, Guyatt G, et al. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. [Updated October 2013]. Available from http://gdt.guidelinedevelopment.org/app/handbook/handbook.html.

9. Santesso N, Carrasco-Labra A, Langendam M, et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. *J Clin Epidemiol* 2016

10. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.

11. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons 2008:187-241.

12. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org/.

13. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

14. Higgins JPT, Savović J, Page MJ, et al. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0), Version 20 October 2016. Available from http://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/rob2-0/ [accessed 19 September 2017].

15. Higgins JPT, Sterne JAC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials. *Cochrane Methods Cochrane Database of Systematic Reviews* 2016;10(Suppl 1):29-31.

16. Hooijmans CR, Rovers MM, de Vries RBM, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.

17. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013;66(4):408-14.

18. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.

19. Dwan K, Gamble C, Kolamunnage-Dona R, et al. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials* 2010;11:52.

20. Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic reviews* 2014;3(1):82.

21. Stewart GB, Higgins JP, Schunemann H, et al. The use of Bayesian networks to assess the quality of evidence from research synthesis: 1. *PLoS One* 2015;10(3):e0114497.

22. Reid EK, Tejani AM, Huan LN, et al. Managing the incidence of selective reporting bias: a survey of Cochrane review groups. *Systematic reviews* 2015;4:85.

23. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ* 2014;349:g6501.

24. Salanti G, Giovane CD, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9(7):e99682.

25. Higgins JP, Del Giovane C, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *Value Health* 2014;17(7):A324.

26. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.

27. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* 2012;65(2):163-78.

28. Viswanathan M, Berkman ND, Dryden DM, et al. AHRQ Methods for Effective Health Care. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or

Exposures: Further Development of the RTI Item Bank. Rockville (MD): Agency for Healthcare Research and Quality (US) 2013.

Table S5. General characteristics of studies evaluating the measurement properties of tools for assessing risk of reporting biases

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Armijo-Olivo 2012[1]	Cochrane risk of bias tool for randomized trials (2008 version)	None	20 trials included in a SR exploring knowledge transfer interventions for cancer pain management.	Cancer pain	None	20	NA	Range 1987-2007	2
Armijo-Olivo 2014[2]	Cochrane risk of bias tool for randomized trials (2011 version)	Inter-rater reliability	Trials of physical therapy interventions included in meta-analyses of a continuous outcome.	Physical therapy for musculoskeletal, cardiorespiratory, neurological or gynaecological conditions	None	109	NA	Not reported	2
Bilandzic 2016[3]	ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool	Inter-rater reliability	Studies included in two SRs of NRSI of the relationship between the use of TZDs and COX-2 inhibitors and major cardiovascular events.	Cardiovascular disease	None	37	NA	Range 2000-2010	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Downs 1998[4]	Downs-Black tool	None	10 randomised controlled trials and 10 non-randomised trials/prospective cohort studies randomly selected from studies identified during a SR of surgery for stress incontinence	Stress incontinence	None	20	NA	Not reported	2
Hartling 2009[5]	Cochrane risk of bias tool for randomized trials (2008 version)	Inter-rater reliability	A convenience sample of 163 randomized trial in child health, which were presented at the annual scientific meetings of the Society for Pediatric Research between 1992 and 1995.	Child health	None	163	NA	Not reported	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Hartling 2011[6]	Cochrane risk of bias tool for randomized trials (2008 version)	Inter-rater reliability	Trials included in a systematic review of long-acting beta agonists (LABA) combined with inhaled corticosteroids (ICS) for adults with persistent asthma.	Asthma	None	107	NA	Median 2004, IQR 2001-2006	2
Hartling 2012[7 8]	Cochrane risk of bias tool for randomized trials (2011 version)	Inter-rater reliability	A sample of 154 trial was randomly selected from among 616 trials published in December 2006 that were previously examined for quality of reporting.	Varied	None	154	NA	All 2006	2
Hayden 2013[9]	QUIPS (Quality In Prognosis Studies) tool	Inter-rater reliability	Studies included in a systematic review of troponin-based risk stratification of patients with	Pulmonary embolism	None	31	NA	Not reported	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
			acute non-massive pulmonary embolism.						
Hoojimans 2014[10]	SYRCLE's RoB tool (SYstematic Review Centre for Laboratory animal Experimentation)	Inter-rater reliability	1 systematic review including 32 papers (no other details provided).	Animal studies (not specified)	None	32	NA	Not reported	2
Jordan 2017[11]	Cochrane risk of bias tool for randomized trials (2011 version)	Inter-rater reliability	Any study that had been included more than once in SRs present on the Cochrane Database of Systematic Reviews in the area of subfertility.	Subfertility	None	28	NA	Not reported	2
Kim 2013[12]	RoBANS (Risk of Bias Assessment Tool for Nonrandomized Studies)	Inter-rater reliability	39 NRSs from four systematic reviews (one by the National Evidence-based Healthcare Collaborating Agency and three Cochrane reviews).	Depression, myocardial infarction, post-partum hemorrhage, chronic non-cancer pain	None	39	NA	Not reported	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Kumar 2016[13]	GRADE	None	10 key questions that were systematically reviewed for a clinical practice guideline for the use of prophylactic vs. therapeutic platelet transfusion in patients with thrombocytopenia.	Thrombocytopenia	10	None	All 2015	NA	18
Llewellyn 2015[14]	SAQAT (Semi-Automated Quality Assessment Tool)	Inter-rater reliability	29 meta-analyses from a purposive sample of SRs of RCTs from the Database of Systematic Reviews of Effects (DARE), and a purposive sample of 15 recent Cochrane reviews in mental health.	Varied	44	None	2006-2013	NA	2
Mustafa 2013[15]	GRADE	None	4 well-conducted and well-reported Cochrane reviews,	Alcohol dependence, asthma,	16	None	2004-2012	NA	4

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
			based on assessment using the AMSTAR tool.	cardiopulmonary bypass					
Norris 2012[16]	ORBIT-I (Outcome Reporting Bias In Trials) classification system for benefit outcomes	Inter-rater reliability; Time to complete assessments	Studies included in three AHRQ-funded comparative effectiveness reviews of randomised trials with drug-drug or drug-placebo comparisons, examining benefit outcomes.	Varied	None	40	NA	2005-2010	2
O'Connor 2015[17]	Downs-Black tool	None	20 studies included in an updated SR which examined the effects of an exercise intervention for chronic musculoskeletal pain.	Chronic musculoskeletal pain	None	20	NA	1997-2008	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Vale 2013[18]	Cochrane risk of bias tool for randomized trials (2011 version)	Agreement between assessments performed using published article only versus published article and data collected during the IPD process.	13 completed IPD meta-analyses of treatments for cancer. Trials had to be published either in full or as an abstract, and a copy of the trial protocol or forms detailing trial design completed by trialists (or both) had to be available.	Cancer pain	None	95	NA	Not reported	2

NA = Not applicable; SR = systematic review

References

1. Armijo-Olivo S, Stiles CR, Hagen NA, et al. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract* 2012;18(1):12-8.

2. Armijo-Olivo S, Ospina M, da Costa BR, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. *PLoS One* 2014;9(5):e96920.

3. Bilandzic A, Fitzpatrick T, Rosella L, et al. Risk of Bias in Systematic Reviews of Non-Randomized Studies of Adverse Cardiovascular Effects of Thiazolidinediones and Cyclooxygenase-2 Inhibitors: Application of a New Cochrane Risk of Bias Tool. *PLoS Med* 2016;13(4):e1001987.

4. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.

5. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.

6. Hartling L, Bond K, Vandermeer B, et al. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. *PLoS One* 2011;6(2):e17242.

7. Hartling L, Hamm M, Milne A, et al. AHRQ Methods for Effective Health Care. Validity and Inter-Rater Reliability Testing of Quality Assessment Instruments. Rockville (MD): Agency for Healthcare Research and Quality (US) 2012.

8. Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol* 2013;66(9):973-81.

9. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.

10. Hooijmans CR, Rovers MM, de Vries RBM, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.

11. Jordan VM, Lensen SF, Farquhar CM. There were large discrepancies in risk of bias tool judgments when a randomized controlled trial appeared in more than one systematic review. *J Clin Epidemiol* 2017;81:72-76.

12. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013;66(4):408-14.

13. Kumar A, Miladinovic B, Guyatt GH, et al. GRADE guidelines system is reproducible when instructions are clearly operationalized even among the guidelines panel members with limited experience with GRADE. *J Clin Epidemiol* 2016;75:115-8.

14. Llewellyn A, Whittington C, Stewart G, et al. The Use of Bayesian Networks to Assess the Quality of Evidence from Research Synthesis: 2. Inter-Rater Reliability and Comparison with Standard GRADE Assessment. *PLoS One* 2015;10(12):e0123511.

15. Mustafa RA, Santesso N, Brozek J, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol* 2013;66(7):736-42; quiz 42.e1-5.

16. Norris SL, Holmer HK, Ogden LA, et al. AHRQ Methods for Effective Health Care. Selective Outcome Reporting as a Source of Bias in Reviews of Comparative Effectiveness. Rockville (MD): Agency for Healthcare Research and Quality (US) 2012.

17. O'Connor SR, Tully MA, Ryan B, et al. Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. *BMC Res Notes* 2015;8:224.

18. Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. *BMJ* 2013;346:f1798.

BMJ Open

Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019703.R1
Article Type:	Research
Date Submitted by the Author:	15-Nov-2017
Complete List of Authors:	Page, Matthew; Monash University, School of Public Health and Preventive Medicine McKenzie, Joanne; Monash University, School of Public Health and Preventive Medicine Higgins, Julian; University of Bristol, Population Health Sciences, Bristol Medical School
<b>Primary Subject Heading</b>:	Research methods
Secondary Subject Heading:	Evidence based practice, Research methods
Keywords:	Publication Bias, Bias (Epidemiology), Review Literature As Topic, Checklist

Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review

Matthew J Page[1,2], Joanne E McKenzie[1], Julian PT Higgins[2]

1. School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

2. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

Correspondence to: Dr. Matthew Page, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne VIC 3004, Australia. Phone: +61 3 9903 0248. Email address: matthew.page@monash.edu

WORD COUNT: 4,129

ABSTRACT

BACKGROUND: Several scales, checklists and domain-based tools for assessing risk of reporting biases exist, but it is unclear how much they vary in content and guidance. We conducted a systematic review of the content and measurement properties of such tools.

METHODS: We searched for potentially relevant articles in Ovid MEDLINE, Ovid EMBASE, Ovid PsycINFO, and Google Scholar from inception to February 2017. One author screened all titles, abstracts and full text articles, and collected data on tool characteristics.

RESULTS: We identified 18 tools that include an assessment of the risk of reporting bias. Tools varied in regard to the type of reporting bias assessed (e.g. bias due to selective publication, bias due to selective non-reporting), and the level of assessment (e.g. for the study as a whole, a particular result within a study, or a particular synthesis of studies). Various criteria are used across tools to designate a synthesis as being at “high” risk of bias due to selective publication (e.g. evidence of funnel plot asymmetry, use of non-comprehensive searches). However, the relative weight assigned to each criterion in the overall judgement is unclear for most of these tools. Tools for assessing risk of bias due to selective non-reporting guide users to assess a study, or an outcome within a study, as “high” risk of bias if no results are reported for an outcome. However, assessing the corresponding risk of bias in a synthesis that is missing the non-reported outcomes is outside the scope of most of these tools. Inter-rater agreement estimates were available for five tools.

CONCLUSION: There are several limitations of existing tools for assessing risk of reporting biases, in terms of their scope, guidance for reaching risk of bias judgements, and measurement properties. Development and evaluation of a new, comprehensive tool, could help overcome present limitations.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- Tools for assessing risk of reporting biases, and studies evaluating their measurement properties, were identified by searching several relevant databases using a search string developed in conjunction with an information specialist.
- Detailed information on the content and measurement properties of existing tools was collected, providing readers with pertinent information to help decide which tools to use in evidence syntheses.
- Screening of articles and data collection were performed by one author only, so it is possible that some relevant articles were missed, or that errors in data collection were made.
- The search of grey literature was not comprehensive, so it is possible that there are other tools for assessing risk of reporting biases, and unpublished studies evaluating measurement properties, that were omitted from this review.

BACKGROUND

The credibility of evidence syntheses can be compromised by reporting biases, which arise when dissemination of research findings is influenced by the nature of the results[1]. For example, there may be bias due to selective publication, where a study is only published if the findings are considered interesting (also known as publication bias)[2]. In addition, bias due to selective non-reporting may occur, where findings (e.g. estimates of intervention efficacy or an association between exposure and outcome) that are statistically non-significant are not reported or are partially reported in a paper (e.g. stating only that "P>0.05")[3]. Alternatively, there may be bias in selection of the reported result, where authors perform multiple analyses for a particular outcome/association, yet only report the result which yielded the most favourable effect estimate[4]. Evidence from cohorts of clinical trials followed from inception suggest that biased dissemination is common. Specifically, on average, half of all trials are not published[1 5], trials with statistically significant results are twice as likely to be published[5], and a third of trials have outcomes that are omitted, added or modified between protocol and publication[6].

Audits of systematic review conduct suggest that most systematic reviewers do not assess risk of reporting biases[7-10]. For example, in a cross-sectional study of 300 systematic reviews indexed in MEDLINE® in February 2014[7], the risk of bias due to selective publication was not considered in 56% of reviews. A common reason for not doing so was that the small number of included studies, or inability to perform a meta-analysis, precluded the use of funnel plots. Only 19% of reviews included a search of a trial registry to identify completed but unpublished trials or pre-specified but non-reported outcomes, and only 7% included a search of another source of data disseminated outside of journal articles. The risk of bias due to selective non-reporting in the included studies was assessed in only 24% of reviews[7]. Another study showed that authors of Cochrane reviews routinely record whether any outcomes that were measured were not reported in the included trials, yet rarely consider if such non-reporting could have biased the results of a synthesis[11].

Previous researchers have summarised the characteristics of tools designed to assess various sources of bias in randomized trials[12-14], non-randomized studies of interventions (NRSI)[14 15], diagnostic test accuracy studies[16], and systematic reviews[14 17]. Others have summarised the performance of statistical methods developed to detect or adjust for reporting biases[18-20]. However, no prior review has focused specifically on tools (i.e. structured instruments such as scales, checklists, or domain-based tools) for assessing the risk of reporting biases. A particular challenge when assessing risk of reporting biases is that existing tools vary in their level of assessment. For example, tools for assessing risk of bias due to selective publication direct assessments at the level of the synthesis, whereas tools for assessing risk of bias due to selective non-reporting within studies can direct assessments at the level of the individual study, at the level of the synthesis, or at both levels. It is unclear how many tools are available to assess different types of reporting bias, and what level they direct assessments at. It is also unclear whether criteria for reaching risk of bias judgements are consistent across existing tools. Therefore, the aim of this research was to conduct a systematic review of the content and measurement properties of such tools.

METHODS

Protocol

Methods for this systematic review were pre-specified in a protocol, which was uploaded to the Open Science Framework in February 2017 (https://osf.io/9ea22/).

Eligibility criteria

Papers were included if the authors described a tool that was designed for use by individuals performing evidence syntheses to assess risk of reporting biases in the included studies or in their synthesis of studies. Tools could assess any type of reporting bias, including bias due to selective publication, bias due to selective non-reporting, or bias in selection of the reported result. Tools could assess the risk of reporting biases in any type of study (e.g. randomized trial of intervention, diagnostic test accuracy study, observational study estimating prevalence of an exposure), and in any type of result (e.g. estimate of intervention efficacy or harm, estimate of diagnostic accuracy, association between exposure and outcome). Eligible tools could take any form, including scales, checklists, and domain-based tools. To be considered a scale, each item had to have a numeric score attached to it, so that an overall summary score could be calculated[12]. To be considered a checklist, the tool had to include multiple questions, but the developers' intention was not to attach a numerical score to each response, or to calculate an overall score[13]. Domain-based tools were those that required users to judge risk of bias or quality within specific domains, and to record the information on which each judgement was based[21].

Tools with a broad scope, for example, to assess multiple sources of bias or the overall quality of the body of evidence, were eligible if one of the items covered risk of reporting bias. Multi-dimensional tools with a statistical component were also eligible (e.g. those that require users to respond to a set of questions about the comprehensiveness of the search, as well as to perform statistical tests for funnel plot asymmetry). In addition, any studies that evaluated the measurement properties of existing tools (e.g. construct validity, inter-rater agreement, time taken to complete assessments) were eligible for inclusion. Papers were eligible regardless of the date or format of publication, but were limited to those written in English.

The following were ineligible:

- articles or book chapters providing guidance on how to address reporting biases, but which do not include a structured tool that can be applied by users (e.g. the 2011 Cochrane Handbook chapter on reporting biases[22]);
- tools developed or modified for use in one particular systematic review;

- tools designed to appraise published systematic reviews, such as the ROBIS tool[23] or AMSTAR[24];
- articles that focus on the development or evaluation of statistical methods to detect or adjust for reporting biases, as these have been reviewed elsewhere[18-20].

Search methods

On 9 February 2017, one author (MJP) searched for potentially relevant records in Ovid MEDLINE (January 1946 to February 2017), Ovid EMBASE (January 1980 to February 2017), and Ovid PsycINFO (January 1806 to February 2017). The search strategies included terms relating to reporting bias, which were combined with a search string used previously by Whiting et al. to identify risk of bias/quality assessment tools[17] (see full Boolean search strategies in online supplementary table S1).

To capture any tools not published by formal academic publishers, we searched Google Scholar using the phrase "reporting bias tool OR risk of bias". One author (MJP) screened the titles of the first 300 records, as recommended by Haddaway et al.[25]. To capture any papers that may have been missed by all searches, one author (MJP) screened the references of included articles. In April 2017, the same author emailed the list of included tools to 15 individuals with expertise in reporting biases and risk of bias assessment, and asked if they were aware of any other tools we had not identified.

Study selection and data collection

One author (MJP) screened all titles and abstracts retrieved by the searches. The same author screened any full text articles retrieved. One author (MJP) collected data from included papers using a standardised data collection form. The following data on included tools were collected:

- type of tool (scale, checklist, or domain-based tool);
- types of reporting bias addressed by the tool;

- level of assessment (i.e. whether users direct assessments at the synthesis or at the individual studies included in the synthesis);
- whether the tool is designed for general use (generic) or targets specific study designs or topic areas (specific);
- items included in the tool;
- how items within the tool are rated;
- methods used to develop the tool (e.g. Delphi study, expert consensus meeting);
- availability of guidance to assist with completion of the tool (e.g. guidance manual).

The following data from studies evaluating measurement properties of an included tool were collected:

- tool evaluated;
- measurement properties evaluated (e.g. inter-rater agreement);
- number of syntheses/studies evaluated;
- publication year of syntheses/studies evaluated;
- areas of health care addressed by syntheses/studies evaluated;
- number of assessors;
- estimate (and precision) of psychometric statistics (e.g. weighted kappa).

Data analysis

We summarised the characteristics of included tools in tables. We calculated the median (interquartile range (IQR)) number of items across all tools, and tabulated the frequency of different criteria used in tools to denote a judgement of "high" risk of reporting bias. We summarised estimates of psychometric statistics, such as weighted kappa to estimate inter-rater agreement[26], by reporting the range of values across studies. For studies reporting weighted kappa, we categorised agreement according to the system proposed by Landis et al.[27], as poor (0.00), slight (0.01-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80), or almost perfect (0.81-1.00).

RESULTS

In total, 5,554 records were identified from the searches, of which we retrieved 165 for full text screening (Figure 1). The inclusion criteria were met by 42 reports summarising 18 tools (Table 1) and 17 studies evaluating the measurement properties of tools[3 4 21 28-66]. A list of excluded papers is presented in online supplementary Table S2. No additional tools were identified by the 15 experts contacted.

Table 1. List of included tools

Article ID	Tool	Scope of tool	Types of reporting biases assessed			Level of assessment[a]
			Selective publication	Selective non-reporting	Selection of the reported result	
Balshem 2013[28]	AHRQ outcome and analysis reporting bias framework	Reporting bias only		✓	✓	Specific outcome/ result in a study
Berkman 2013[29]	AHRQ tool for evaluating the risk of reporting bias	Reporting bias only	✓	✓		Specific synthesis of studies
Downes 2016[30]	AXIS tool (Appraisal tool for Cross-Sectional Studies)	Multiple sources of bias		✓		Study
Downs 1998[31]	Downs-Black tool	Multiple sources of bias			✓	Study
Guyatt 2011[33-37]	GRADE	Multiple sources of bias	✓	✓		Specific synthesis of studies
Hayden 2013[38]	QUIPS (Quality In Prognosis Studies) tool	Multiple sources of bias		✓		Study
Higgins 2008[21 39 40]	Cochrane risk of bias tool for randomized trials (RoB 1.0)	Multiple sources of bias		✓	✓	Study
Higgins 2016[41 42]	RoB 2.0 revised tool for assessing risk of bias in randomized trials	Multiple sources of bias			✓	Specific result in a study
Hoojimans 2014[43]	SYRCLE's RoB tool (SYstematic Review Centre for Laboratory animal Experimentation)	Multiple sources of bias		✓	✓	Study
Kim 2013[44]	RoBANS (Risk of Bias Assessment Tool for Nonrandomized Studies)	Multiple sources of bias		✓	✓	Study
Kirkham	ORBIT-I (Outcome Reporting Bias In Trials)	Reporting bias		✓		Specific outcome

Article ID	Tool	Scope of tool	Types of reporting biases assessed			Level of assessment[a]
			Selective publication	Selective non-reporting	Selection of the reported result	
2010[3 32]	classification system for benefit outcomes	only				in a study
Meader 2014[45 46]	SAQAT (Semi-Automated Quality Assessment Tool)	Multiple sources of bias	✓	✓		Specific synthesis of studies
Reid 2015[47]	Selective reporting bias algorithm	Reporting bias only		✓	✓	Study
Saini 2014[48]	ORBIT-II (Outcome Reporting Bias In Trials) classification system for harm outcomes	Reporting bias only		✓		Specific outcome/ result in a study
Salanti 2014[49 50]	Framework for evaluating the quality of evidence from a network meta-analysis	Multiple sources of bias	✓	✓		Specific synthesis of studies
Sterne 2016[4]	ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool	Multiple sources of bias			✓	Specific result in a study
Viswanathan 2012[51]	RTI Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures	Multiple sources of bias		✓		Study
Viswanathan 2013[52]	RTI Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures	Multiple sources of bias		✓		Study

[a] Level of assessment classified as: "study" when assessments are directed at a study as a whole (e.g. tool used to assess whether *any* outcomes in a study were not reported); "specific outcome/result in a study" when assessments are directed at a specific outcome or result within a study (e.g. tools used to assess whether a particular outcome, such as pain, was not reported) or; "specific synthesis of studies" when assessments are directed at a specific synthesis (e.g. tool used to assess whether a particular synthesis, such as a meta-analysis of pain, is missing unpublished studies).

General characteristics of included tools

Nearly all of the included tools (16/18 [89%]) were domain-based, where users judge risk of bias or quality within specific domains (Table 2; individual characteristics of each tool are presented in online supplementary Table S3). All tools were designed for generic rather than specific use. Five tools focused solely on the risk of reporting biases[3 28 29 47 48]; the remainder addressed reporting biases and other sources of bias/methodological quality (e.g. problems with randomization, lack of blinding). Half of the tools (9/18 [50%]) addressed only one type of reporting bias (e.g. bias due to selective non-reporting only). Tools varied in regard to the study design that they assessed (i.e. randomized trial, non-randomized study of an intervention, laboratory animal experiment). The publication year of the tools ranged from 1998 to 2016 (the earliest was the Downs-Black tool[31], a 27-item tool assessing multiple sources of bias, one of which focuses on risk of bias in selection of the reported result).

Assessments for half of the tools (9/18 [50%]) are directed at an individual study (e.g. tool is used to assess whether *any outcomes in a study* were not reported). In 5/18 (28%) tools, assessments are directed at a specific outcome or result within a study (e.g. tool is used to assess whether *a particular outcome in a study*, such as pain, was not reported). In a few tools (4/18 [22%]), assessments are directed at a specific synthesis (e.g. tool is used to assess whether *a particular synthesis*, such as a meta-analysis of studies examining pain as an outcome, is missing unpublished studies).

The content of the included tools was informed by various sources of data. The most common included a literature review of items used in existing tools or a literature review of empirical evidence of bias (9/18 [50%]), ideas generated at an expert consensus meeting (8/18 [44%]) and pilot feedback on a preliminary version of the tool (7/18 [39%]). The most common type of guidance available for the tools was a brief annotation per item/response option (9/18 [50%]). A detailed guidance manual is available for four (22%) tools.

Table 2. Summary of general characteristics of included tools

Characteristic	**Summary data (n = 18 tools)**
Type of tool	
Domain-based	16 (89%)
Checklist	1 (6%)
Scale	1 (6%)
Scope of tool	
Assessment of reporting bias only	5 (28%)
Assessment of multiple sources of bias/quality	13 (72%)
Types of reporting bias assessed	
Bias due to selective publication only	0 (0%)
Bias due to selective non-reporting only	6 (33%)
Bias in selection of the reported result only	3 (17%)
Bias due to selective publication and bias due to selective non-reporting	4 (22%)
Bias due to selective non-reporting and bias in selection of the reported result	5 (28%)
Total number of items in the tool	7 (5-13)
Number of items relevant to risk of reporting bias	1 (1-2)
Number of response options for risk of reporting bias judgement	3 (3-3)
Types of study designs to which the tool applies	
Randomized trials only	5 (28%)
Systematic reviews only	3 (17%)
Non-randomized studies of interventions only	2 (11%)
Randomized trials and non-randomized studies of interventions	2 (11%)
Non-randomized studies of interventions or exposures	2 (11%)
Other (cross-sectional studies, animal studies, network meta-analyses, prognosis studies)	4 (22%)
Level of assessment of risk of reporting bias	
Study as a whole	9 (50%)
Specific outcome/result in a study	5 (28%)
Specific synthesis of studies	4 (22%)
Data sources used to inform tool content[a]	
Literature review (e.g. of items in existing tools, or empirical evidence)	9 (50%)
Ideas generated at expert consensus meeting	8 (44%)
Pilot feedback on preliminary version of the tool	7 (39%)

Characteristic	Summary data (n = 18 tools)
Data from psychometric or cognitive testing[b]	5 (28%)
Other (e.g. adaptation of existing tool)	5 (28%)
Delphi study responses	2 (11%)
No methods stated	2 (11%)
Guidance available	
Brief annotation per item/response option	9 (50%)
Detailed guidance manual	4 (22%)
Worked example for each response option	2 (11%)
Detailed annotation per item/response option	1 (6%)
None	2 (11%)

Summary data given as number (percent) or median (IQR).

[a] The percentages in this category do not sum to 100% since the development of some tools was informed by multiple data sources.

[b] Psychometric testing includes any evaluation of the measurement properties (e.g. construct validity, inter-rater reliability, test-retest reliability) of a draft version of the tool. Cognitive testing includes use of qualitative methods (e.g. interview) to explore whether assessors who are using the tool for the first time were interpreting the tool and guidance as intended.

Tool content

Four tools include items for assessing risk of bias due to both selective publication and selective non-reporting[29 33 45 49]. One of these tools (the AHRQ tool for evaluating the risk of reporting bias[29]) directs users to assess a particular synthesis, where a single risk of bias judgement is made based on information about unpublished studies and underreported outcomes. In the other three tools (the GRADE framework, and two others which are based on GRADE[33 45 49]), the different sources of reporting bias are assessed in separate domains (bias due to selective non-reporting is considered in a “study limitations (risk of bias)” domain, while bias due to selective publication is considered in a “publication bias” domain).

Five tools[21 28 43 44 47] guide users to assess risk of bias due to both selective non-reporting and selection of the reported result (that is, problems with outcomes/results that *are not* reported and those that *are* reported, respectively). Four of these tools, which include the Cochrane risk of bias tool for randomized trials[21] and three others which are based on the Cochrane tool[43 44 47], direct assessments at the study level. That is, a whole study is rated at “high” risk of reporting bias if *any* outcome/result in the study has been omitted, or fully reported, on the basis of the findings.

Some of the tools designed to assess the risk of bias due to selective non-reporting ask users to assess, for particular outcomes of interest, whether the outcome was not reported or only partially reported in the study on the basis of its results (e.g. ORBIT tools[3 48], the AHRQ outcome reporting bias framework[28], and GRADE[34]). This allows users to perform multiple outcome-level assessments of the risk of reporting bias (rather than one assessment for the study as a whole). In total, 15 tools include a mechanism for assessing risk of bias due to selective non-reporting in studies, but assessing the corresponding risk of bias in a synthesis that is missing the non-reported outcomes is not within the scope of 11 of these tools [3 21 28 30 38 43 44 47 48 51 52].

A variety of criteria are used in existing tools to inform a judgement of “high” risk of bias due to selective publication (Table 3), selective non-reporting (Table 4), and selection of the reported result (Table 5) (more detail is provided in online supplementary Table S4). In the four tools with an assessment of risk of bias due to selective publication, “high” risk criteria include evidence of funnel plot asymmetry, discrepancies between published and unpublished studies, use of non-comprehensive searches, and presence of small, “positive” studies with for-profit interest (Table 3). However, not all of these criteria appear in all tools (only evidence of funnel plot asymmetry does), and the relative weight assigned to each criterion in the overall risk of reporting bias judgement is clear for only one tool (the Semi-Automated Quality Assessment Tool (SAQAT)[45 46]).

All 15 tools with an assessment of the risk of bias due to selective non-reporting suggest that the risk of bias is “high” when it is clear that an outcome was measured but no results were reported (Table

4). Fewer of these tools (n=8 [53%]) also recommend a “high” risk judgement when results for an outcome are partially reported (e.g. it is stated that the result was non-significant, but no effect estimate or summary statistics are presented).

The eight tools that include an assessment of the risk of bias in selection of the reported result recommend various criteria for a “high” risk judgement (Table 5). These include when some outcomes that were not pre-specified are added post-hoc (in 4 [50%] tools), or when it is likely that the reported result for a particular outcome has been selected, on the basis of the findings, from amongst multiple outcome measurements or analyses within the outcome domain (in 2 [25%] tools).

Table 3. Criteria used in existing tools to inform a judgement of "high" risk of bias due to selective publication

"High" risk of bias criteria proposed in existing tools	AHRQ RRB	GRADE	SAQAT	NMA-Quality	Total n (%)
Assessment directed at a specific synthesis (e.g. meta-analysis)					
Evidence of funnel plot asymmetry (based on visual inspection of funnel plot or statistical test for funnel plot asymmetry)	✓	✓	✓	✓	4 (100)
Smaller studies tend to demonstrate more favourable results (based on visual assessment, without funnel plot)	✓				1 (25)
Clinical decision would differ for estimates from a fixed-effect versus a random-effects model, because the findings from a fixed-effect model are closer to the null	✓				1 (25)
Substantial heterogeneity in the meta-analysis cannot be explained by some clinical or methodological factor	✓				1 (25)
At least one study is affected by non-publication or non-accessibility	✓				1 (25)
Presence of small (often "positive") studies with for-profit interest in the synthesis		✓		✓	2 (50)
Presence of early studies (i.e. set of small, "positive" trials addressing a novel therapy) in the synthesis		✓		✓	2 (50)
Discrepancy in findings between published and unpublished trials		✓	✓	✓	3 (75)
Search strategies were not comprehensive		✓	✓	✓	3 (75)
Methods to identify all available evidence were not comprehensive		✓		✓	2 (50)
Grey literature were not searched			✓		1 (25)
Restrictions to study selection on the basis of language were applied			✓		1 (25)
Industry influence may apply to studies included in the synthesis			✓		1 (25)

AHRQ RRB = AHRQ tool for evaluating the risk of reporting bias[29]; GRADE = GRADE rating of quality of evidence[34-37]; NMA-Quality = Framework for evaluating the quality of evidence from a network meta-analysis[49]; SAQAT = Semi-Automated Quality Assessment Tool[45 46].

Table 4. Criteria used in existing tools to inform a judgement of "high" risk of bias due to selective non-reporting

"High" risk of bias criteria proposed in existing tools	AHRQ ORB	AHRQ RRB	AXIS	GRADE	QUIPS	RoB 1.0	SYRCLE RoB	RoBANS	ORBIT-I	SAQAT	Reid	ORBIT-II	NMA-Quality	RTI 2012	RTI 2013	Total n (%)
Assessment directed at study as a whole																
One or more outcomes of interest were clearly measured, but no results were reported			✓		✓	✓	✓	✓			✓			✓	✓	8 (53)
One or more outcomes of interest are reported incompletely so that they cannot be entered in a meta-analysis						✓		✓								2 (13)
The study report fails to include results for a key outcome that would be expected to have been reported for such a study						✓	✓	✓						✓	✓	5 (33)
Assessment directed at a specific outcome																
Particular outcome clearly measured, but no results were reported	✓	✓		✓					✓			✓	✓			6 (40)
Particular outcome of interest is reported incompletely so that it cannot be entered in a meta-analysis (typically stating only that P>0.05).	✓	✓		✓					✓			✓	✓			6 (40)

"High" risk of bias criteria proposed in existing tools	AHRQ ORB	AHRQ RRB	AXIS	GRADE	QUIPS	RoB 1.0	SYRCLE RoB	RoBANS	ORBIT-I	SAQAT	Reid	ORBIT-II	NMA-Quality	RTI 2012	RTI 2013	Total n (%)
Judgment says particular outcome is likely to have been measured and analysed but not reported on the basis of its results	✓	✓		✓					✓			✓	✓			6 (40)
Composite outcomes are presented without the individual component outcomes				✓												1 (7)
Data were not reported consistently for the outcome of interest										✓						1 (7)
Assessment directed at a specific synthesis																
Selective non-reporting suspected in a number of included studies		✓		✓						✓			✓			4 (27)

AHRQ ORB = AHRQ outcome and analysis reporting bias framework[28]; AHRQ RRB = AHRQ tool for evaluating the risk of reporting bias[29]; AXIS = Appraisal tool for Cross-Sectional Studies[30]; GRADE = GRADE rating of quality of evidence[34-37]; NMA-Quality = Framework for evaluating the quality of evidence from a network meta-analysis[49]; ORBIT-I = Outcome Reporting Bias In Trials classification system for benefit outcomes[3 32]; ORBIT-II = Outcome Reporting Bias In Trials classification system for harm outcomes[48]; QUIPS = Quality In Prognosis Studies tool[38]; Reid = Reid et al. selective reporting bias algorithm[47]; RoB 1.0 = Cochrane risk of bias tool for randomized trials[21 39 40]; RoBANS = Risk of Bias Assessment Tool for Nonrandomized Studies[44]; RTI 2012 = RTI Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures[51]; RTI 2013 = RTI Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures[52]; SAQAT = Semi-Automated Quality Assessment Tool[45 46]; SYRCLE RoB = SYstematic Review Centre for Laboratory animal Experimentation risk of bias tool[43].

Table 5. Criteria used in existing tools to inform a judgement of "high" risk of bias in selection of the reported result

"High" risk of bias criteria proposed in existing tools	AHRQ ORB	Downs-Black	RoB 1.0	RoB 2.0	SYRCLE RoB	RoBANS	Reid	ROBINS-I	Total n (%)
Assessment directed at study as a whole									
One or more reported outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse event)			✓		✓	✓	✓		4 (50)
One or more outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified			✓		✓				2 (15)
One or more retrospective, unplanned, subgroup analysis was reported		✓							1 (13)
Any analyses that had not been planned at the outset of the study were not clearly indicated		✓							1 (13)
Assessment directed at a specific outcome/result									
Particular outcome was not pre-specified but results were reported	✓								1 (13)
Reported result for a particular outcome is likely to have been selected, on the basis of the findings, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain				✓				✓	2 (25)
Reported result for a particular outcome is likely to have been selected, on the basis of the findings, from multiple analyses of the data				✓				✓	2 (25)
Reported result for a particular outcome is likely to have been selected, on the basis of the findings, from different subgroups								✓	1 (13)

AHRQ ORB = AHRQ outcome and analysis reporting bias framework[28]; Downs-Black = Downs-Black tool[31]; Reid = Reid et al. selective reporting bias algorithm[47]; RoB 1.0 = Cochrane risk of bias tool for randomized trials[21 39 40]; RoB 2.0 = Revised tool for assessing risk of bias in randomized trials[41 42]; RoBANS = Risk of Bias Assessment Tool for Nonrandomized Studies[44]; ROBINS-I = Risk Of Bias In Non-randomized Studies of Interventions tool[4]; SYRCLE RoB = SYstematic Review Centre for Laboratory animal Experimentation risk of bias tool[43].

General characteristics of studies evaluating measurement properties of included tools

Despite identifying 17 studies that evaluated measurement properties of an included tool, psychometric statistics for the risk of reporting bias component were available only from 12 studies[43 44 54-60 62 64 66] (the other five studies include only data on properties of the multi-dimensional tool as a whole[31 53 61 63 65]) (online supplementary Table S5). Nearly all 12 studies (11 [92%]) evaluated inter-rater agreement between two assessors; eight of these studies reported weighted kappa (κ) values, but only two described the weighting scheme[55 62]. Eleven studies[43 44 54-60 64 66] evaluated the measurement properties of tools for assessing risk of bias in a study due to selective non-reporting or risk of bias in selection of the reported result; in these 11 studies, a median of 40 (IQR 32-109) studies were assessed. One study[62] evaluated a tool for assessing risk of bias in a synthesis due to selective publication, in which 44 syntheses were assessed. In the studies evaluating inter-rater agreement, all involved two assessors.

Results of evaluation studies

Five studies[54 56-58 60] included data on the inter-rater agreement of assessments of risk of bias due to selective non-reporting using the Cochrane risk of bias tool for randomized trials[21] (Table 6). Weighted kappa (κ) values in four studies[54 56-58] ranged from 0.13 to 0.50 (sample size ranged from 87 to 163 studies), suggesting slight to moderate agreement[27]. In the other study[60], the percent agreement in selective non-reporting assessments in trials that were included in two different Cochrane reviews was low (43% of judgements were in agreement). Two other studies found that inter-rater agreement of selective non-reporting assessments were substantial for SYRCLE's RoB tool (κ = 0.62, n = 32)[43], but poor for the RoBANS tool (κ = 0, n = 39)[44]. There was substantial agreement between raters in the assessment of risk of bias due to selective publication using the SAQAT (κ = 0.63, n = 29)[62]. The inter-rater agreement of assessments of risk of bias in selection of the reported result using the ROBINS-I tool[4] was moderate for NRSI included in a review of the effect of cyclooxygenase-2 (COX-2) inhibitors on cardiovascular events (κ = 0.45, n = 21), and substantial for

NRSI included in a review of the effect of thiazolidinediones on cardiovascular events (κ = 0.78, n = 16)[55].

Table 6. Reported measurement properties of tools with an assessment of the risk of reporting bias

Study ID	Tool	Measurement property	Sample size	Areas of health care addressed	Weighted kappa (95% CI)	Weighting scheme	Interpretation of kappa[a]
Armijo-Olivo 2014[54]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between two external reviewers)	87	Musculoskeletal, cardiorespiratory, neurological, and gynaecological conditions	0.5 (CI not reported)	Not described	Moderate agreement
Armijo-Olivo 2014[54]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between two external reviewers and Cochrane reviewers)	87	See above	0.13 (CI not reported)	Not described	Slight agreement
Hartling 2009[56]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting	163	Child health	0.13 (95% CI -0.05 to 0.31)	Not described	Slight agreement
Hartling 2011[57]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting	107	Asthma	0.4 (95% CI 0.14 to 0.67)	Not described	Fair agreement
Hartling 2012[58] [59]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between two reviewers, all trials)	124	Varied	0.27 (95% CI 0.06 to 0.49)	Not described	Fair agreement
Hartling 2012[58] [59]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between pairs of reviewers across different centres, all trials)	30	Varied	0.08 (95% CI -0.09 to 0.26)	Not described	Slight agreement
Jordan 2017[60]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between judgements of trials appearing in two SRs)	28	Subfertility	Not reported[b]	Not applicable	Not applicable
Vale 2013[66]	RoB 1.0	Agreement between selective non-reporting assessments performed using published article only versus published article and data collected during the individual participant data process	95	Cancer pain	Not reported[b]	Not applicable	Not applicable
Hoojimans	SYRCLE RoB	Inter-rater agreement of assessments of risk of	32	Animal studies (not	0.62 (CI not	Not	Substantial

Study ID	Tool	Measurement property	Sample size	Areas of health care addressed	Weighted kappa (95% CI)	Weighting scheme	Interpretation of kappa[a]
2014[43]		bias due to selective non-reporting		specified)	reported)	described	agreement
Kim 2013[44]	RoBANS	Inter-rater agreement of assessments of risk of bias due to selective non-reporting	39	Depression, myocardial infarction, post-partum hemorrhage, chronic non-cancer pain	0 (CI not reported)	Not described	Poor agreement
Llewellyn 2015[62]	SAQAT	Inter-rater agreement of assessments of risk of bias due to selective publication (between two SAQAT raters)	29	Varied	0.63 (95% CI 0.17 to 1)	Quadratic	Substantial agreement
Llewellyn 2015[62]	SAQAT	Inter-rater agreement of assessments of risk of bias due to selective publication (between one rater using SAQAT and one using the standard GRADE approach)	15	Varied	Not reported[b]	Not applicable	Not applicable
Norris 2012[64]	ORBIT-I	Inter-rater agreement of ORBIT-I classifications of risk of bias due to selective non-reporting	40	Varied	Not calculated, as too little variation in judgements	Not applicable	Not applicable
Bilandzic 2016[55]	ROBINS-I	Inter-rater agreement of assessments of risk of bias in selection of the reported result	16	Thiazolidinediones and cardiovascular events	0.78 (CI not reported)	Linear	Substantial agreement
Bilandzic 2016[55]	ROBINS-I	Inter-rater agreement of assessments of risk of bias in selection of the reported result	21	COX-2 inhibitors and cardiovascular events	0.45 (CI not reported)	Linear	Moderate agreement

[a] Interpretation of kappa based on categorisation system defined by Landis et al.[27]. [b] Data presented as percent agreement, not weighted kappa. ORBIT-I = Outcome Reporting Bias In Trials classification system for benefit outcomes[3 32]; RoB 1.0 = Cochrane risk of bias tool for randomized trials[21 39 40]; RoBANS = Risk of Bias Assessment Tool for Nonrandomized Studies[44]; ROBINS-I = Risk Of Bias In Non-randomized Studies of Interventions tool[4]; SAQAT = Semi-Automated Quality Assessment Tool[45 46]; SRs = systematic reviews; SYRCLE RoB = SYstematic Review Centre for Laboratory animal Experimentation risk of bias tool[43].

DISCUSSION

From a systematic search of the literature, we identified 18 tools designed for use by individuals performing evidence syntheses to assess risk of reporting biases in the included studies or in their synthesis of studies. The tools varied with regard to the type of reporting bias assessed (e.g. bias due to selective publication, bias due to selective non-reporting), and the level of assessment (e.g. for the study as a whole, a particular outcome within a study, or a particular synthesis of studies). Various criteria are used across tools to designate a synthesis as being at "high" risk of bias due to selective publication (e.g. evidence of funnel plot asymmetry, use of non-comprehensive searches). However, the relative weight assigned to each criterion in the overall judgement is not clear for most of these tools. Tools for assessing risk of bias due to selective non-reporting guide users to assess a study, or an outcome within a study, as "high" risk of bias if no results are reported for an outcome. However, assessing the corresponding risk of bias in a synthesis that is missing the non-reported outcomes is outside the scope of most of these tools. Inter-rater agreement estimates were available for five tools[4 21 43 44 62], and ranged from poor to substantial; however the sample sizes of most evaluations were small, and few described the weighting scheme used to calculate kappa.

Strengths and limitations

There are several strengths of this research. Methods were conducted in accordance with a systematic review protocol (https://osf.io/9ea22/). Published articles were identified by searching several relevant databases using a search string developed in conjunction with an information specialist[17], and by contacting experts to identify tools missed by the search. Detailed information on the content and measurement properties of existing tools was collected, providing readers with pertinent information to help decide which tools to use in future reviews. However, the findings need to be considered in light of some limitations. Screening of articles and data collection were performed by one author only. It is therefore possible that some relevant articles were missed, or that errors in data collection were made. The search for unpublished tools was not comprehensive (only Google Scholar was searched), so it is possible that other tools for assessing risk of reporting biases exist. Further, restricting the search to articles in English was done to expedite the review process, but may have resulted in loss of information about tools written in other languages, and additional evidence on measurement properties of tools.

Comparison with other studies

Other systematic reviews of risk of bias tools[12-17] have restricted inclusion to tools developed for particular study designs (e.g. randomized trials, diagnostic test accuracy studies), where the authors recorded all the sources of bias addressed. A different approach was taken in the current review, where all tools (regardless of study design) that address a particular source of bias were examined. By focusing on one source of bias only, the analysis of included items and criteria for risk of bias judgements was more detailed than that recorded previously. Some of the existing reviews of tools[15] considered tools that were developed or modified in the context of a specific systematic review. However, such tools were excluded from the current review as they are unlikely to have been developed systematically[15 67], and are difficult to find (all systematic reviews conducted during a particular period would need to have been examined for the search to be considered exhaustive).

Explanations and implications

Of the 18 tools identified, only four (22%) included a mechanism for assessing risk of bias due to selective publication, which is the type of reporting bias that has been investigated by methodologists most often[2]. This is perhaps unsurprising given that hundreds of statistical methods to "detect" or "adjust" for bias due to selective publication have been developed[18]. These statistical methods may be considered by methodologists and systematic reviewers as the tools of choice for assessing this type of bias. However, application of these statistical methods without considering other factors (e.g. existence of registered but unpublished studies, vested interests of investigators) is not sufficiently comprehensive, and could lead to incorrect conclusions about the risk of bias due to selective publication. Further, there are many limitations of these statistical approaches, in terms of their underlying assumptions, statistical power, which is often low because most meta-analyses include few studies[7], and the need for specialist statistical software to apply them[19 68]. These factors may have limited their use in practice, and potentially explain why a large number of systematic reviewers currently ignore the risk of bias due to selective publication[7-9 69].

Our analysis suggests that the factors that need to be considered to assess risk of reporting biases adequately (e.g. comprehensiveness of the search, amount of data missing from the synthesis due to unpublished studies and underreported outcomes) are fragmented. A similar problem was occurring a decade ago with the assessment of risk of bias in randomized trials. Some authors assessed only problems with randomization, while others focused on whether trials were not “double blinded”, or had any missing participant data[70]. It was not until all the important bias domains were brought together into a structured, domain-based tool to assess the risk of bias in randomized trials[21], that systematic reviewers started to consider risk of bias in trials comprehensively. A similar initiative to link all the components needed to judge the risk of reporting biases into a comprehensive new tool may improve the credibility of evidence syntheses.

In particular, there is an emergent need for a new tool to assess the risk that a synthesis is affected by reporting biases. This tool could guide users to consider risk of bias in a synthesis due to both selective publication and selective non-reporting, given that both practices lead to the same consequence: evidence missing from the synthesis[11]. Such a tool would complement recently developed tools for assessing risk of bias within studies (RoB 2.0[41] and ROBINS-I[4]) which include a domain for assessing the risk of bias in selection of the reported result, but no mechanism to assess risk of bias due to selective non-reporting. Careful thought would need to be given as to how to weigh up various pieces of information underpinning the risk of bias judgement. For example, users will need guidance on how evidence of known, unpublished studies (as identified from trial registries, protocols or regulatory documents) should be considered alongside evidence that is more speculative (e.g. funnel plots suggesting that studies may be missing). Further, guidance for the tool will need to emphasise the value of seeking documents other than published journal articles (e.g. protocols) to inform risk of bias judgements. Preparation of a detailed guidance manual may enhance the usability of the tool, minimise misinterpretation and increase reliability in assessments. Once developed, evaluations of the measurement properties of the tool, such as inter-rater agreement and construct validity, should be conducted to explore whether modifications to the tool are necessary.

Conclusions

There are several limitations of existing tools for assessing risk of reporting biases in studies or syntheses of studies, in terms of their scope, guidance for reaching risk of bias judgements, and measurement properties. Development and evaluation of a new, comprehensive tool, could help overcome present limitations.

Acknowledgments

Not applicable.

Competing Interests

We have read the journal's policy and have the following competing interests: JPTH led or participated in the development of four of the included tools (the current Cochrane risk of bias tool for randomized trials, the RoB 2.0 tool for assessing risk of bias in randomized trials, the ROBINS-I tool for assessing risk of bias in non-randomized studies of interventions, and the framework for assessing quality of evidence from a network meta-analysis). MJP participated in the development of one of the included tools (the RoB 2.0 tool for assessing risk of bias in randomized trials). All authors are participating in the development of a new tool for assessing risk of reporting biases in systematic reviews.

Author Contributions

MJP conceived and designed the study, collected data, analysed the data, and wrote the first draft of the article. JM and JPTH provided input on the study design and contributed to revisions of the article. All authors approved the final version of the submitted article.

Data sharing statement

The study protocol, data collection form, and the raw data and statistical analysis code for this study are available on the Open Science Framework: https://osf.io/3jdaa/

Funding

There was no direct funding for this study. MJP is supported by an Australian National Health and Medical Research Council (NHMRC) Early Career Fellowship (1088535). JEM is supported by a NHMRC Australian Public Health Fellowship (1072366). JPTH is funded in part by Cancer Research UK Programme Grant C18281/A19169; is a member of the MRC Integrative Epidemiology Unit at the University of Bristol, which is supported by the UK Medical Research Council and the University of Bristol (grant MC_UU_12013/9); and is a member of the MRC ConDuCT-II Hub (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures; grant MR/K025643/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Chan A-W, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *The Lancet* 2014;383(9913):257-66.

2. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010;14:8.

3. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.

4. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.

5. Schmucker C, Schell LK, Portalupi S, et al. Extent of non-publication in cohorts of studies approved by research ethics committees or included in trial registries. *PLoS One* 2014;9(12):e114023.

6. Jones CW, Keil LG, Holland WC, et al. Comparison of registered and published outcomes in randomized controlled trials: a systematic review. *BMC Med* 2015;13:282.

7. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med* 2016;13(5):e1002028.

8. Koletsi D, Valla K, Fleming PS, et al. Assessment of publication bias required improvement in oral health systematic reviews. *J Clin Epidemiol* 2016;76:118-24

9. Hedin RJ, Umberham BA, Detweiler BN, et al. Publication Bias and Nonreporting Found in Majority of Systematic Reviews and Meta-analyses in Anesthesiology Journals. *Anesth Analg* 2016;123(4):1018-25.

10. Ziai H, Zhang R, Chan AW, et al. Search for unpublished data by systematic reviewers: an audit. *BMJ open* 2017;7(10):e017737.

11. Page MJ, Higgins JPT. Rethinking the assessment of risk of bias due to selective reporting: a cross-sectional study. *Systematic reviews* 2016;5(1):108.

12. Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;16(1):62-73.

13. Armijo Olivo S, Macedo LG, Gadotti IC, et al. Scales to assess the quality of randomized controlled trials: a systematic review. *Phys Ther* 2008;88(2):156-75.

14. Bai A, Shukla VK, Bak G, et al. Quality Assessment Tools Project Report. Ottawa: Canadian Agency for Drugs and Technologies in Health, 2012.

15. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007;36(3):666-76.

16. Whiting P, Rutjes AW, Dinnes J, et al. A systematic review finds that diagnostic reviews fail to incorporate quality despite available tools. *J Clin Epidemiol* 2005;58(1):1-12.

17. Whiting P, Davies P, Savovic J, et al. Evidence to inform the development of ROBIS, a new tool to assess the risk of bias in systematic reviews, September 2013. Available from https://www.researchgate.net/publication/303312018_Evidence_to_inform_the_development_of_ROBIS_a_new_tool_to_assess_the_risk_of_bias_in_systematic_reviews [accessed 1 August 2017].

18. Mueller KF, Meerpohl JJ, Briel M, et al. Methods for detecting, quantifying and adjusting for dissemination bias in meta-analysis are described. *J Clin Epidemiol* 2016;80:25-33.

19. Jin ZC, Zhou XH, He J. Statistical methods for dealing with publication bias in meta-analysis. *Stat Med* 2015;34(2):343-60.

20. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002.

21. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

22. Sterne JAC, Egger M, Moher D. Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions Version 510 [updated March 2011] 2011.

23. Whiting P, Savovic J, Higgins JP, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;69:225-34.

24. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.

25. Haddaway NR, Collins AM, Coughlin D, et al. The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. *PLoS One* 2015;10(9):e0138237.

26. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20(1):37-46.

27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-74.

28. Balshem H, Stevens A, Ansari M, et al. Finding grey literature evidence and assessing for outcome and analysis reporting biases when comparing medical interventions: AHRQ and the Effective Health Care Program. (Prepared by the Oregon Health and Science University and the University of Ottawa Evidence-based Practice Centers under Contract Nos. 290-2007-10057-I and 290-2007-10059-I.) AHRQ Publication No. 13(14)-EHC096-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

29. Berkman ND, Lohr KN, Ansari M, et al. Chapter 15 Appendix A: A Tool for Evaluating the Risk of Reporting Bias (in Chapter 15: Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update). Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I). AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm

30. Downes MJ, Brennan ML, Williams HC, et al. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ open* 2016;6:e011458.

31. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.

32. Dwan K, Gamble C, Kolamunnage-Dona R, et al. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials* 2010;11:52.

33. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.

34. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol* 2011;64(4):407-15.

35. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64(12):1277-82.

36. Schünemann H, Brożek J, Guyatt G, et al. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. [Updated October 2013]. Available from http://gdt.guidelinedevelopment.org/app/handbook/handbook.html.

37. Santesso N, Carrasco-Labra A, Langendam M, et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. *J Clin Epidemiol* 2016

38. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.

39. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons 2008:187-241.

40. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org/.

41. Higgins JPT, Savović J, Page MJ, et al. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0), Version 20 October 2016. Available from https://sites.google.com/site/riskofbiastool/ [accessed 19 September 2017].

42. Higgins JPT, Sterne JAC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials. *Cochrane Methods Cochrane Database of Systematic Reviews* 2016;10(Suppl 1):29-31.

43. Hooijmans CR, Rovers MM, de Vries RBM, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.

44. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013;66(4):408-14.

45. Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic reviews* 2014;3(1):82.

46. Stewart GB, Higgins JP, Schunemann H, et al. The use of Bayesian networks to assess the quality of evidence from research synthesis: 1. *PLoS One* 2015;10(3):e0114497.

47. Reid EK, Tejani AM, Huan LN, et al. Managing the incidence of selective reporting bias: a survey of Cochrane review groups. *Systematic reviews* 2015;4:85.

48. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ* 2014;349:g6501.

49. Salanti G, Giovane CD, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9(7):e99682.

50. Higgins JP, Del Giovane C, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *Value Health* 2014;17(7):A324.

51. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* 2012;65(2):163-78.

52. Viswanathan M, Berkman ND, Dryden DM, et al. AHRQ Methods for Effective Health Care. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or

Exposures: Further Development of the RTI Item Bank. Rockville (MD): Agency for Healthcare Research and Quality (US) 2013.

53. Armijo-Olivo S, Stiles CR, Hagen NA, et al. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract* 2012;18(1):12-8.

54. Armijo-Olivo S, Ospina M, da Costa BR, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. *PLoS One* 2014;9(5):e96920.

55. Bilandzic A, Fitzpatrick T, Rosella L, et al. Risk of Bias in Systematic Reviews of Non-Randomized Studies of Adverse Cardiovascular Effects of Thiazolidinediones and Cyclooxygenase-2 Inhibitors: Application of a New Cochrane Risk of Bias Tool. *PLoS Med* 2016;13(4):e1001987.

56. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.

57. Hartling L, Bond K, Vandermeer B, et al. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. *PLoS One* 2011;6(2):e17242.

58. Hartling L, Hamm M, Milne A, et al. AHRQ Methods for Effective Health Care. Validity and Inter-Rater Reliability Testing of Quality Assessment Instruments. Rockville (MD): Agency for Healthcare Research and Quality (US) 2012.

59. Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol* 2013;66(9):973-81.

60. Jordan VM, Lensen SF, Farquhar CM. There were large discrepancies in risk of bias tool judgments when a randomized controlled trial appeared in more than one systematic review. *J Clin Epidemiol* 2017;81:72-76.

61. Kumar A, Miladinovic B, Guyatt GH, et al. GRADE guidelines system is reproducible when instructions are clearly operationalized even among the guidelines panel members with limited experience with GRADE. *J Clin Epidemiol* 2016;75:115-8.

62. Llewellyn A, Whittington C, Stewart G, et al. The Use of Bayesian Networks to Assess the Quality of Evidence from Research Synthesis: 2. Inter-Rater Reliability and Comparison with Standard GRADE Assessment. *PLoS One* 2015;10(12):e0123511.

63. Mustafa RA, Santesso N, Brozek J, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol* 2013;66(7):736-42; quiz 42.e1-5.

64. Norris SL, Holmer HK, Ogden LA, et al. AHRQ Methods for Effective Health Care. Selective Outcome Reporting as a Source of Bias in Reviews of Comparative Effectiveness. Rockville (MD): Agency for Healthcare Research and Quality (US) 2012.

65. O'Connor SR, Tully MA, Ryan B, et al. Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. *BMC Res Notes* 2015;8:224.

66. Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. *BMJ* 2013;346:f1798.

67. Whiting PF, Rutjes AW, Westwood ME, et al. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol* 2013;66(10):1093-104.

68. Sterne JAC, Egger M, Moher D, et al. Chapter 10: Addressing reporting biases. In: Higgins JPT, Churchill R, Chandler J, et al., eds. Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0. (updated June 2017). Available from www.training.cochrane.org/handbook: Cochrane 2017.

69. Atakpo P, Vassar M. Publication bias in dermatology systematic reviews and meta-analyses. *J Dermatol Sci* 2016;82(2):69-74.

70. Lundh A, Gotzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. *BMC Med Res Methodol* 2008;8:22.

Figure legends

Figure 1. Flow diagram of identification, screening and inclusion of studies. [a]Refers to records identified from Ovid MEDLINE, Ovid EMBASE, Ovid PsycINFO, and Google Scholar. [b]Refers to records identified from screening references of included articles.

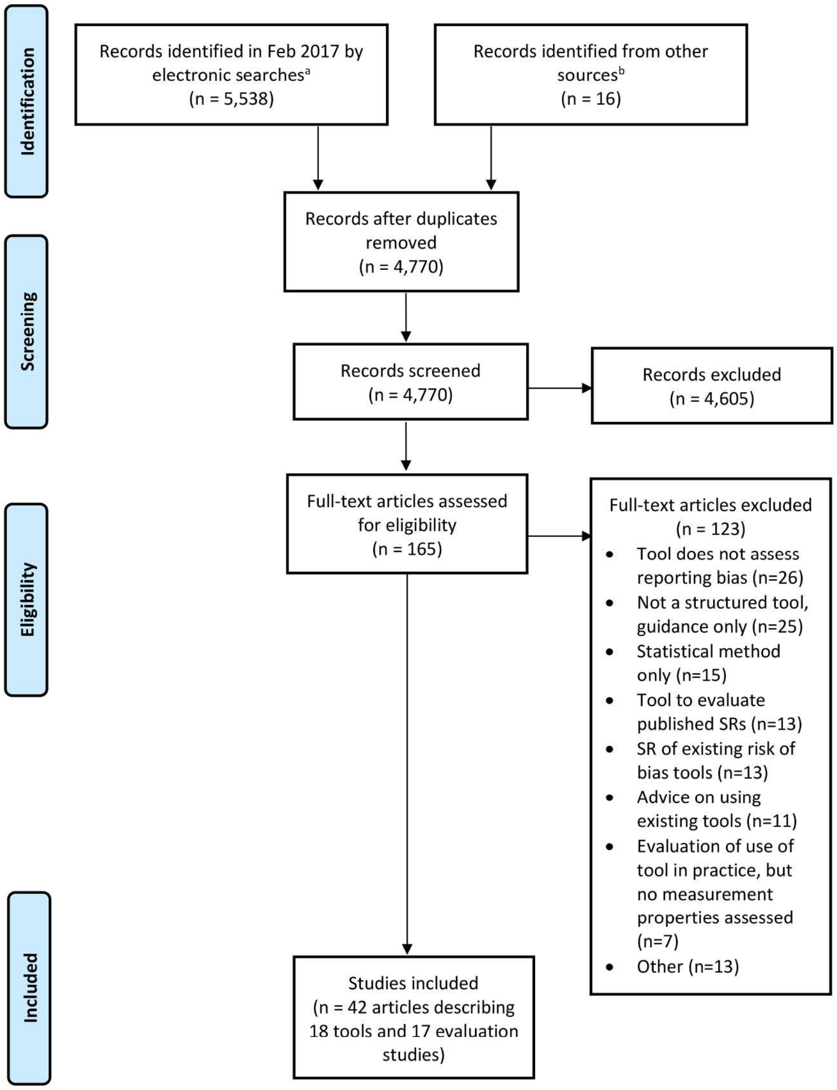
Identification
Screening
Eligibility
Included
Records identified in Feb 2017 by electronic searches[a] (n = 5,538)
Records identified from other sources[b] (n = 16)
Records after duplicates removed (n = 4,770)
Records screened (n = 4,770)
Records excluded (n = 4,605)
Full-text articles assessed for eligibility (n = 165)
Full-text articles excluded (n = 123)
• Tool does not assess reporting bias (n=26)
• Not a structured tool, guidance only (n=25)
• Statistical method only (n=15)
• Tool to evaluate published SRs (n=13)
• SR of existing risk of bias tools (n=13)
• Advice on using existing tools (n=11)
• Evaluation of use of tool in practice, but no measurement properties assessed (n=7)
• Other (n=13)
Studies included (n = 42 articles describing 18 tools and 17 evaluation studies)

Figure 1. Flow diagram of identification, screening and inclusion of studies. aRefers to records identified from Ovid MEDLINE, Ovid EMBASE, Ovid PsycINFO, and Google Scholar. bRefers to records identified from screening references of included articles.

171x238mm (300 x 300 DPI)

Table S1. Search strategies

```
Database: Ovid MEDLINE(R) <1946 to 9 February 2017>
Search Strategy:
--------------------------------------------------------------------------------
1     ((tool or tools or instrument$ or checklist$ or check list$ or scale or scales) and (quality or
      methodolog$ or method or methods)).ti.
2     (quality adj10 (score or scores or scoring or rating or rate) adj5 (methodolog$ or method or
      methods)).tw.
3     (guideline$ and (quality or methodolog$ or method or methods)).ti.
4     ((assess$ or apprais$ or critical$) adj3 (systematic review$ or meta-analys$ or
      metaanalys$)).ti.
5     ((score or scores or scoring or rating or rate) and (quality or methodolog$ or method or
      methods)).ti.
6     ((quality or methodology) adj3 (review or meta-analys$ or metaanalys$) adj3 (assess$ or
      method$)).tw.
7     (quality adj3 article$).tw.
8     (critical$ adj2 (apprais$ or evaluat$)).tw.
9     ((apprais$ or evaluat$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
10    (guideline$ adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
11    or/1-10
12    Checklist/
13    11 or 12
14    Publication Bias/
15    exp "bias (epidemiology)"/
16    (bias adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
17    ((quality or bias or methodolog$) adj3 (systematic review$ or meta-analys$ or
      metaanalys$)).tw.
18    (bias$ adj3 (publication$ or disseminat$ or language$ or reporting or grey or gray or citation$
      or time delay or time lag or conference or abstract)).tw.
19    or/14-18
20    13 and 19



Database: Embase <1980 to 2017 Week 06>
Search Strategy:
--------------------------------------------------------------------------------
1     "Review Literature as Topic"/
2     "meta analysis (topic)"/
3     meta analysis/
4     "systematic review (topic)"/
5     systematic review/
6     systematic review$.tw.
7     (meta-analys$ or metaanalys$).tw.
8     or/1-7
9     (bias adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
10    ((quality or bias or methodolog$) adj3 (systematic review$ or meta-analys$ or
      metaanalys$)).tw.
11    (bias$ adj3 (publication$ or disseminat$ or language$ or reporting or grey or gray or citation$
      or time delay or time lag or conference or abstract)).tw.
```

12 "internal validity"/
13 publishing/
14 or/9-13
15 ((tool or tools or instrument$ or checklist$ or check list$ or scale or scales) and (quality or methodolog$ or method or methods)).ti.
16 (quality adj10 (score or scores or scoring or rating or rate) adj5 (methodolog$ or method or methods)).tw.
17 (guideline$ and (quality or methodolog$ or method or methods)).ti.
18 ((assess$ or apprais$ or critical$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).ti.
19 ((score or scores or scoring or rating or rate) and (quality or methodolog$ or method or methods)).ti.
20 ((quality or methodology) adj3 (review or meta-analys$ or metaanalys$) adj3 (assess$ or method$)).tw.
21 (quality adj3 article$).tw.
22 (critical$ adj2 (apprais$ or evaluat$)).tw.
23 ((apprais$ or evaluat$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
24 (guideline$ adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
25 or/15-24
26 checklist/
27 25 or 26
28 8 and 14 and 27
29 limit 28 to embase

Database: PsycINFO <1806 to February Week 1 2017>
Search Strategy:
----
1 meta-analysis/
2 systematic review$.tw.
3 (meta-analys$ or metaanalys$).tw.
4 or/1-3
5 ((tool or tools or instrument$ or checklist$ or check list$ or scale or scales) and (quality or methodolog$ or method or methods)).ti.
6 (quality adj10 (score or scores or scoring or rating or rate) adj5 (methodolog$ or method or methods)).tw.
7 (guideline$ and (quality or methodolog$ or method or methods)).ti.
8 ((assess$ or apprais$ or critical$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).ti.
9 ((score or scores or scoring or rating or rate) and (quality or methodolog$ or method or
1. methods)).ti.
10 ((quality or methodology) adj3 (review or meta-analys$ or metaanalys$) adj3 (assess$ or method$)).tw.
11 (quality adj3 article$).tw.
12 (critical$ adj2 (apprais$ or evaluat$)).tw.
13 ((apprais$ or evaluat$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
14 (guideline$ adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
15 checklist/
16 or/5-15
17 (bias adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.

18 ((quality or bias or methodolog$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.

19 (bias$ adj3 (publication$ or disseminat$ or language$ or reporting or grey or gray or citation$ or time delay or time lag or conference or abstract)).tw.

20 bias.mp.

21 or/17-20

22 4 and 16 and 21

Table S2. Excluded studies

Reference	**Reason for exclusion**
Armijo-Olivo S, Cummings GG, Fuentes J, Saltaji H, Ha C, Chisholm A, et al. Identifying items to assess methodological quality in physical therapy trials: a factor analysis. Physical Therapy 2014;94(9):1272-84.	Paper does not report on a structured tool
Armijo-Olivo S, Fuentes J, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: a descriptive analysis. BMC Medical Research Methodology 2013;13:116.	Systematic review of tools
Armijo-Olivo S, Fuentes J, Rogers T, Hartling L, Saltaji H, Cummings GG. How should we evaluate the risk of bias of physical therapy trials?: a psychometric and meta-epidemiological approach towards developing guidelines for the design, conduct, and reporting of RCTs in Physical Therapy (PT) area: a study protocol. Syst Rev 2013;2:88.	Protocol for development of new tool
Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. International Journal of Evidence-Based Healthcare 2015;13(3):132-40.	Refers to a tool to assess quality of published systematic reviews
Arrive L, Renard R, Carrat F, Belkacem A, Dahan H, Le Hir P, et al. A scale of methodological quality for clinical studies of radiologic examinations. Radiology 2000;217(1):69-74.	Tool does not assess reporting bias
Atakpo P, Vassar M. Publication bias in dermatology systematic reviews and meta-analyses. Journal of Dermatological Science 2016;82(2):69-74.	Describes statistical methods only
Ballard M, Montgomery P. Risk of bias in overviews of reviews: a scoping review of methodological guidance and four-item checklist. Research Synthesis Methods 2017;8(1):92-108.	Refers to a tool to assess quality of published systematic reviews
Balzer K. Assessing the quality of research needs to go beyond scoring: Commentary on Crowe and Sheppard (2011). International Journal of Nursing Studies 2012;49(8):1048-50.	Commentary
Bartlett WA, Braga F, Carobene A, Coskun A, Prusa R, Fernandez-Calle P, et al. A checklist for critical appraisal of studies of biological variation. Clinical Chemistry and Laboratory Medicine 2015;53(6):879-85.	Tool does not assess reporting bias
Bashir R, Dunn AG. Systematic review protocol assessing the processes for linking clinical trial registries and their published results. BMJ Open 2016;6(10):e013048.	Paper does not report on a structured tool
Beck NB, Becker RA, Boobis A, Fergusson D, Fowle JR, Goodman J, et al. Instruments for assessing risk of bias and other methodological criteria of animal studies: omission of well-established methods. Environmental Health Perspectives 2014;122(3):A66-7.	Commentary
Berkman ND, Lohr KN, Morgan LC, Kuo T-M, Morton SC. Interrater	Tool does not assess

Reference	Reason for exclusion
reliability of grading strength of evidence varies with the complexity of the evidence in systematic reviews. Journal of Clinical Epidemiology 2013;66(10):1105-17.e1.	reporting bias
Burda BU, Holmer HK, Norris SL. Limitations of A Measurement Tool to Assess Systematic Reviews (AMSTAR) and suggestions for improvement. Systematic Reviews 2016;5:58.	Refers to a tool to assess quality of published systematic reviews
Cartes-Velasquez RA, Manterola C, Aravena P, Moraga J. Reliability and validity of MINCIR scale for methodological quality in dental therapy research. Brazilian Oral Research 2014;28.	Tool does not assess reporting bias
Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Research Synthesis Methods 2012;3(2):161-76.	Describes statistical methods only
da Costa BR, Hilfiker R, Egger M. PEDro's bias: summary quality scores should not be used in meta-analysis. Journal of Clinical Epidemiology 2013;66(1):75-7.	Commentary
Dahm P. Raising the bar for systematic reviews with Assessment of Multiple Systematic Reviews (AMSTAR). BJU International 2017;119(2):193.	Refers to a tool to assess quality of published systematic reviews
Dalton DR, Aguinis H, Dalton CM, Bosco FA, Pierce CA. Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. Personnel Psychology 2012;65(2):221-49.	Paper does not report on a structured tool
David SP, Ware JJ, Chu IM, Loftus PD, Fusar-Poli P, Radua J, et al. Potential reporting bias in fMRI studies of the brain. PloS One 2013;8(7):e70104.	Paper does not report on a structured tool
Davino-Ramaya C, Krause LK, Robbins CW, Harris JS, Koster M, Chan W, et al. Transparency matters: Kaiser Permanente's National Guideline Program methodological processes. The Permanente Journal 2012;16(1):55-62.	Refers to a tool to assess quality of published systematic reviews
Dawson A, Raphael KG, Glaros A, Axelsson S, Arima T, Ernberg M, et al. Development of a quality-assessment tool for experimental bruxism studies: reliability and validity. Journal of Orofacial Pain 2013;27(2):111-22.	Tool does not assess reporting bias
Deshpande S, Misso K, Westwood M, Stirk L, De Kock S, Clayton D, et al. Not all cochrane reviews are good quality systematic reviews. Value in Health 2016;19(7):A371.	Refers to a tool to assess quality of published systematic reviews
Disher T, Benoit B, Johnston C, Campbell-Yeo M. Skin-to-skin contact for procedural pain in neonates: acceptability of novel systematic review synthesis methods and GRADEing of the evidence. Journal of Advanced Nursing 2017;73(2):504-19.	Paper does not report on a structured tool
Dreier M, Borutta B, Stahmeyer J, Krauth C, Walter U. Comparison of tools for assessing the methodological quality of primary and secondary studies in health technology assessment reports in Germany. GMS Health Technology Assessment 2010;6.	Systematic review of tools

Reference	Reason for exclusion
Dreyer N, Velentgas P, Duddy A, Westrich KD, Dubois RW. Grace checklist: Rating the strength of evidence for observational studies of comparative effectiveness. Value in Health 2012;15(4):A5.	Tool does not assess reporting bias
Dreyer NA, Velentgas P, Westrich K, Dubois R. The GRACE checklist for rating the quality of observational studies of comparative effectiveness: a tale of hope and caution. Journal of Managed Care & Specialty Pharmacy 2014;20(3):301-8.	Tool does not assess reporting bias
Dreyer NA, Velentgas P, Westrich K, Dubois RW. GRACE: A validated checklist for identifying robust observational studies of comparative effectiveness. Pharmacoepidemiol Drug Saf 2013;22:356.	Tool does not assess reporting bias
Dreyer NA, Velentgas P, Westrich KD, Dubois RW. There but for grace? a validated screening tool for quality observational studies of comparative effectiveness. Value in Health 2013;16(3):A21.	Tool does not assess reporting bias
Drucker AM, Fleming P, Chan A-W. Research Techniques Made Simple: Assessing Risk of Bias in Systematic Reviews. The Journal of Investigative Dermatology 2016;136(11):e109-e14.	Guidance on using existing tools
Dwan K, Altman DG, Clarke M, Gamble C, Higgins JP, Sterne JA, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. PLoS Med 2014;11(6):e1001666.	Paper does not report on a structured tool
Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. PLoS One 2013;8(7):e66844.	Paper does not report on a structured tool
Dwan K, Kirkham JJ, Williamson PR, Gamble C. Selective reporting of outcomes in randomised controlled trials in systematic reviews of cystic fibrosis. BMJ Open 2013;3(6).	Evaluation of use of tool in practice, but no measurement properties assessed
Fantony JJ, Gopalakrishna A, Noord MV, Inman BA. Reporting Bias Leading to Discordant Venous Thromboembolism Rates in the United States Versus Non-US Countries Following Radical Cystectomy: A Systematic Review and Meta-analysis. European Urology Focus 2016;2(2):189-96.	Paper does not report on a structured tool
Fitzgerald A, Coop C. Validation and modification of the Graphical Appraisal Tool for Epidemiology (GATE) for appraising systematic reviews in evidence-based guideline development. Health Outcomes Research in Medicine 2011;2(1):e51-e9.	Refers to a tool to assess quality of published systematic reviews
Frosi G, Riley RD, Williamson PR, Kirkham JJ. Multivariate meta-analysis helps examine the impact of outcome reporting bias in Cochrane rheumatoid arthritis reviews. J Clin Epidemiol 2015;68(5):542-50.	Evaluation of use of tool in practice, but no measurement properties assessed
Furukawa TA, Miura T, Chaimani A, Leucht S, Cipriani A, Noma H, et al. Using the contribution matrix to evaluate complex study limitations in a network meta-analysis: a case study of bipolar maintenance pharmacotherapy review. BMC Res Notes 2016;9:218.	Describes statistical methods only

Reference	Reason for exclusion
Ghogomu EAT, Maxwell LJ, Buchbinder R, Rader T, Pardo Pardo J, Johnston RV, et al. Updated method guidelines for cochrane musculoskeletal group systematic reviews and metaanalyses. The Journal of Rheumatology 2014;41(2):194-205.	Guidance on using existing tools
Golder S, Loke YK, Bland M. Unpublished data can be of value in systematic reviews of adverse effects: methodological overview. Journal of Clinical Epidemiology 2010;63(10):1071-81.	Paper does not report on a structured tool
Golder S, Loke YK. Is there evidence for biased reporting of published adverse effects data in pharmaceutical industry-funded studies? British Journal of Clinical Pharmacology 2008;66(6):767-73.	Paper does not report on a structured tool
Goodyear-Smith FA, van Driel ML, Arroll B, Del Mar C. Analysis of decisions made in meta-analyses of depression screening and the risk of confirmation bias: a case study. BMC Med Res Methodol 2012;12:76.	Paper does not report on a structured tool
Grant S, Pedersen ER, Osilla KC, Kulesza M, D'Amico EJ. It is time to develop appropriate tools for assessing minimal clinically important differences, performance bias and quality of evidence in reviews of behavioral interventions. Addiction 2016;111(9):1533-5.	Paper does not report on a structured tool
Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostatistics (Oxford, England) 2001;2(4):463-71.	Describes statistical methods only
Haddaway NR, Woodcock P, Macura B, Collins A. Making literature reviews more reliable through application of lessons from systematic reviews. Conservation Biology 2015;29(6):1596-605.	Guidance on using existing tools
Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. Statistics in Medicine 2000;19(24):3325-36.	Describes statistical methods only
Heck NC, Mirabito LA, LeMaire K, Livingston NA, Flentje A. Omitted data in randomized controlled trials for anxiety and depression: A systematic review of the inclusion of sexual orientation and gender identity. Journal of Consulting and Clinical Psychology 2017;85(1):72-6.	Paper does not report on a structured tool
Higgins JPT, Lane PW, Anagnostelis B, Anzures-Cabrera J, Baker NF, Cappelleri JC, et al. A tool to assess the quality of a meta-analysis. Research Synthesis Methods 2013;4(4):351-66.	Refers to a tool to assess quality of published systematic reviews
Hoy D, Brooks P, Woolf A, Blyth F, March L, Bain C, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. J Clin Epidemiol 2012;65(9):934-9.	Tool does not assess reporting bias
Hsu W, Speier W, Taira RK. Automated extraction of reported statistical analyses: towards a logical representation of clinical trial literature. AMIA Annual Symposium proceedings AMIA Symposium 2012;2012:350-9.	Paper does not report on a structured tool
Ioannidis JPA, Munafo MR, Fusar-Poli P, Nosek BA, David SP. Publication and other reporting biases in cognitive sciences:	Paper does not report on a

Reference	Reason for exclusion
detection, prevalence, and prevention. Trends in Cognitive Sciences 2014;18(5):235-41.	structured tool
Ioannidis JPA, Trikalinos TA. An exploratory test for an excess of significant findings. Clinical Trials 2007;4(3):245-53.	Describes statistical methods only
Ioannidis JPA, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. CMAJ 2007;176(8):1091-6.	Describes statistical methods only
Jarde A, Losilla J-M, Vives J, Rodrigo MF. Q-Coh: A tool to screen the methodological quality of cohort studies in systematic reviews and meta-analysis. International Journal of Clinical and Health Psychology 2013;13(2):138-46.	Tool does not assess reporting bias
Jefferson T, Jones MA, Doshi P, Del Mar CB, Hama R, Thompson MJ, et al. Risk of bias in industry-funded oseltamivir trials: comparison of core reports versus full clinical study reports. BMJ Open 2014;4(9):e005253.	Evaluation of use of tool in practice, but no measurement properties assessed
Johnson BT, Low RE, MacDonald HV. Panning for the gold in health research: incorporating studies' methodological quality in meta-analysis. Psychology & Health 2015;30(1):135-52.	Describes statistical methods only
Johnston BC, Patrick DL, Busse JW, Schunemann HJ, Agarwal A, Guyatt GH. Patient-reported outcomes in meta-analyses--Part 1: assessing risk of bias and combining outcomes. Health and Quality of Life Outcomes 2013;11:109.	Guidance on using existing tools
Jorgensen L, Paludan-Muller AS, Laursen DR, Savovic J, Boutron I, Sterne JA, et al. Evaluation of the Cochrane tool for assessing risk of bias in randomized clinical trials: overview of published comments and analysis of user practice in Cochrane and non-Cochrane reviews. Syst Rev 2016;5:80.	Evaluation of use of tool in practice, but no measurement properties assessed
Jurgens T, Whelan AM, MacDonald M, Lord L. Development and evaluation of an instrument for the critical appraisal of randomized controlled trials of natural products. BMC Complement Altern Med 2009;9:11.	Tool does not assess reporting bias
Jurgens TM, Whelan AM. Development and evaluation of an instrument for the critical appraisal of randomized controlled trials of natural products. Canadian Journal of Hospital Pharmacy 2011;64(1):68.	Tool does not assess reporting bias
Katikireddi SV, Egan M, Petticrew M. How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. Journal of Epidemiology and Community Health 2015;69(2):189-95.	Audit of tools used in systematic reviews
Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar S, Grimmer KA. A systematic review of the content of critical appraisal tools. BMC Med Res Methodol 2004;4:22.	Systematic review of tools
Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in	Describes statistical methods only

Reference	Reason for exclusion
systematic reviews. Statistics in Medicine 2012;31(20):2179-95.	
Kocsis JH, Gerber AJ, Milrod B, Roose SP, Barber J, Thase ME, et al. A new scale for assessing the quality of randomized clinical trials of psychotherapy. Comprehensive Psychiatry 2010;51(3):319-24.	Tool does not assess reporting bias
Kovacs FM, Abraira V. Language Bias in a Systematic Review of Chronic Pain: How to Prevent the Omission of Non-English Publications? The Clinical Journal of Pain 2004;20(3):199-200.	Paper does not report on a structured tool
Krauth D, Woodruff TJ, Bero L. Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. Environmental Health Perspectives 2013;121(9):985-92.	Systematic review of tools
Kromrey JD, Rendina-Gobioff G. On Knowing What We Do Not Know: An Empirical Comparison of Methods to Detect Publication Bias in Meta-Analysis. Educational and Psychological Measurement 2006;66(3):357-73.	Describes statistical methods only
Lamont RF. A quality assessment tool to evaluate tocolytic studies. BJOG 2006;113(Suppl 3):96-9.	Tool does not assess reporting bias
Langendam M, Carrasco-Labra A, Santesso N, Mustafa RA, Brignardello-Petersen R, Ventresca M, et al. Improving GRADE evidence tables part 2: A systematic survey of explanatory notes shows more guidance is needed. J Clin Epidemiol 2016;74:19-27.	Evaluation of use of tool in practice, but no measurement properties assessed
Liebherz S, Schmidt N, Rabung S. How to assess the quality of psychotherapy outcome studies: A systematic review of quality assessment criteria. Psychotherapy Research 2016;26(5):573-89.	Systematic review of tools
Liebherz S, Schmidt N, Rabung S. Study Quality and its Influence on Treatment Outcome in Studies on the Effectiveness of Inpatient Psychotherapy - A Meta-Analysis. PPmP Psychotherapie Psychosomatik Medizinische Psychologie 2016;66(1):31-8.	Not written in English
Lohr KN, Carey TS. Assessing "best evidence": issues in grading the quality of studies for systematic reviews. The Joint Commission Journal on Quality Improvement 1999;25(9):470-9.	Guidance on using existing tools
Lonjon G, Porcher R, Ergina P, Fouet M, Boutron I. Potential Pitfalls of Reporting and Bias in Observational Studies With Propensity Score Analysis Assessing a Surgical Procedure: A Methodological Systematic Review. Ann Surg 2016:no pagination.	Paper does not report on a structured tool
Lundh A, Gotzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. BMC Med Res Methodol 2008;8:22.	Guidance on using existing tools
Lynch HN, Goodman JE, Tabony JA, Rhomberg LR. Systematic comparison of study quality criteria. Regul Toxicol Pharmacol 2016;76:187-98.	Systematic review of tools
Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, et al. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. PLoS Biology 2015;13(10):e1002273.	Tool does not assess reporting bias

Reference	Reason for exclusion
Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003;83(8):713-21.	Tool does not assess reporting bias
Malmivaara A. Methodological considerations of the GRADE method. Annals of Medicine 2015;47(1):1-5.	Guidance on using existing tools
Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. Journal of the American Medical Informatics Association 2016;23(1):193-201.	Model to semi-automate Cochrane risk of bias tool
McDonagh MS, Peterson K, Balshem H, Helfand M. US Food and Drug Administration documents can provide unpublished evidence relevant to systematic reviews. Journal of Clinical Epidemiology 2013;66(10):1071-81.	Paper does not report on a structured tool
McShane BB, Bockenholt U, Hansen KT. Adjusting for Publication Bias in Meta-Analysis: An Evaluation of Selection Methods and Some Cautionary Notes. Perspectives on Psychological Science 2016;11(5):730-49.	Describes statistical methods only
Millard LAC, Flach PA, Higgins JPT. Machine learning to assist risk-of-bias assessments in systematic reviews. International Journal of Epidemiology 2016;45(1):266-77.	Model to semi-automate Cochrane risk of bias tool
Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Controlled Clinical Trials 1995;16(1):62-73.	Systematic review of tools
Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLoS Med 2014;11(10):e1001744.	Refers to a tool to assess quality of published systematic reviews
Moyer A, Finney JW. Rating methodological quality: toward improved assessment and investigation. Accountability in Research 2005;12(4):299-313.	Guidance on using existing tools
Mueller KF, Briel M, Strech D, Meerpohl JJ, Lang B, Motschall E, et al. Dissemination bias in systematic reviews of animal research: a systematic review. PloS One 2014;9(12):e116016.	Paper does not report on a structured tool
Mueller KF, Meerpohl JJ, Briel M, Antes G, von Elm E, Lang B, et al. Detecting, quantifying and adjusting for publication bias in meta-analyses: protocol of a systematic review on methods. Systematic Reviews 2013;2:60.	Describes statistical methods only
Mueller KF, Meerpohl JJ, Briel M, Antes G, von Elm E, Lang B, et al. Methods for detecting, quantifying, and adjusting for dissemination bias in meta-analysis are described. J Clin Epidemiol 2016;80:25-33.	Describes statistical methods only
Nakagawa S, Noble DWA, Senior AM, Lagisz M. Meta-evaluation of meta-analysis: ten appraisal questions for biologists. BMC Biology 2017;15(1):18.	Refers to a tool to assess quality of published systematic reviews
Nolting A, Perleth M, Langer G, Meerpohl JJ, Gartlehner G, Kaminski-	Not written in English

Reference	Reason for exclusion
Hartenthaler A, et al. [GRADE guidelines: 5. Rating the quality of evidence: publication bias]. Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen 2012;106(9):670-6.	
Norris SL, Moher D, Reeves BC, Shea B, Loke Y, Garner S, et al. Issues relating to selective reporting when including non-randomized studies in systematic reviews on the effects of healthcare interventions. Res Synth Methods 2013;4(1):36-47.	Guidance on using existing tools
Nurmatov UB, Xiong T, Kroes MA. Evaluation of quality assessment tools for non-randomised controlled trials assessing surgical interventions: A systematic review of systematic reviews. Value in Health 2015;18(7):A722.	Systematic review of tools
Odierna DH, Forsyth SR, White J, Bero LA. The cycle of bias in health research: a framework and toolbox for critical appraisal training. Accountability in Research 2013;20(2):127-41.	Paper does not report on a structured tool
Palma Perez S, Delgado Rodriguez M. [Practical considerations on detection of publication bias]. Gac Sanit 2006;20(Suppl 3):10-6.	Not written in English
Pearson M, Peters J. Outcome reporting bias in evaluations of public health interventions: evidence of impact and the potential role of a study register. Journal of Epidemiology and Community Health 2012;66(4):286-9.	Evaluation of use of tool in practice, but no measurement properties assessed
Petticrew M, Egan M, Thomson H, Hamilton V, Kunkler R, Roberts H. Publication bias in qualitative research: what becomes of qualitative research presented at conferences? Journal of Epidemiology and Community Health 2008;62(6):552-4.	Paper does not report on a structured tool
Pigott TD, Valentine JC, Polanin JR, Williams RT, Canada DD. Outcome-Reporting Bias in Education Research. Educational Researcher 2013;42(8):424-32.	Paper does not report on a structured tool
Pirracchio R, Resche-Rigon M, Chevret S, Journois D. Do simple screening statistical tools help to detect reporting bias? Annals of Intensive Care 2013;3(1):29.	Describes statistical methods only
Quigley JM, Thompson J, Halfpenny N, Scott DA. Critical appraisal of non-randomized controlled trials-a review of recommended and commonly used tools. Value in Health 2014;17(3):A203.	Systematic review of tools
Quigley JM, Thompson JC, Halfpenny NJ, Scott DA. Critical appraisal of real world evidence-a review of recommended and commonly used tools. Value in Health 2015;18(7):A684.	Systematic review of tools
Quintana DS. From pre-registration to publication: A non-technical primer for conducting a meta-analysis to synthesize correlational data. Front Psychol 2015;6:1549.	Paper does not report on a structured tool
Rangel SJ, Kelsey J, Colby CE, Anderson J, Moss RL. Development of a quality assessment scale for retrospective clinical studies in pediatric surgery. Journal of Pediatric Surgery 2003;38(3):390-6.	Tool does not assess reporting bias
Rosella L, Bowman C, Pach B, Morgan S, Fitzpatrick T, Goel V. The development and validation of a meta-tool for quality appraisal of	Tool does not assess reporting bias

Reference	Reason for exclusion
public health evidence: Meta Quality Appraisal Tool (MetaQAT). Public Health 2016 Jul;136:57-65.	
Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. International Journal of Epidemiology 2007;36(3):666-76.	Systematic review of tools
Santaguida PL, Riley CM, Matchar DB. Chapter 5: Assessing risk of bias as a domain of quality in medical test studies. Journal of General Internal Medicine 2012;27(Suppl 1):S33-S8.	Guidance on using existing tools
Savovic J, Weeks L, Sterne JA, Turner L, Altman DG, Moher D, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. Syst Rev 2014;3:37.	Evaluation of use of tool in practice, but no measurement properties assessed
Seehra J, Pandis N, Koletsi D, Fleming PS. Use of quality assessment tools in systematic reviews was varied and inconsistent. J Clin Epidemiol 2016;69:179-84.e5.	Audit of tools used in systematic reviews
Shamliyan T, Kane RL, Dickinson S. A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. Journal of Clinical Epidemiology 2010;63(10):1061-70.	Systematic review of tools
Shamliyan TA, Kane RL, Ansari MT, Raman G, Berkman ND, Grant M, et al. Development quality criteria to evaluate nontherapeutic studies of incidence, prevalence, or risk factors of chronic diseases: pilot study of new checklists. Journal of Clinical Epidemiology 2011;64(6):637-57.	Tool does not assess reporting bias
Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10.	Refers to a tool to assess quality of published systematic reviews
Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. Journal of Clinical Epidemiology 2009;62(10):1013-20.	Refers to a tool to assess quality of published systematic reviews
Shuang M, Zhao C, Zhang L, Shang HC. Using SYRCLE tools to evaluate the methodological quality of animal experiments of stroke in China. Chinese Journal of Evidence-Based Medicine 2016;16(5):592-7.	Not written in English
Singh S, Khosla S. Suboptimal choice of methodology for meta-analysis and publication bias assessment. The American Journal of Cardiology 2015;115(12):1782-3.	Describes statistical methods only
Smyth RM, Kirkham JJ, Jacoby A, Altman DG, Gamble C, Williamson PR. Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists. BMJ 2011;342:c7153.	Paper does not report on a structured tool
Sohani ZN, Meyre D, de Souza RJ, Joseph PG, Gandhi M, Dennis BB, et al. Assessing the quality of published genetic association studies in	Tool does not assess

Reference	Reason for exclusion
meta-analyses: the quality of genetic studies (Q-Genie) tool. BMC Genet 2015;16:50.	reporting bias
Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technology Assessment (Winchester, England) 2010;14(8):iii-193.	Paper does not report on a structured tool
Spooner CH, Pickard AS, Menon D. Edmonton Quality Assessment Tool for Drug Utilization Reviews: EQUATDUR-2: the development of a scale to assess the methodological quality of a drug utilization review. Medical Care 2000;38(9):948-58.	Tool does not assess reporting bias
Tate RL, Perdices M, Rosenkoetter U, Wakim D, Godbee K, Togher L, et al. Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: the 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. Neuropsychological Rehabilitation 2013;23(5):619-38.	Tool does not assess reporting bias
Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters M, et al. AHRQ Methods for Effective Health Care Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012.	Guidance on using existing tools
Voss PH, Rehfuess EA. Quality appraisal in systematic reviews of public health interventions: an empirical study on the impact of choice of tool on meta-analysis. Journal of Epidemiology and Community Health 2013;67(1):98-104.	Evaluation of existing tools
Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2008. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed 7/03/2017).	Tool does not assess reporting bias
Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J. A systematic review finds that diagnostic reviews fail to incorporate quality despite available tools. Journal of Clinical Epidemiology 2005;58(1):1-12.	Systematic review of tools
Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.	Tool does not assess reporting bias
Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine 2011;155(8):529-36.	Tool does not assess reporting bias
Wiart L, Kolaski K, Vogtle LK, Butler C, Romeiser Logan L, Hickman R, et al. Inter-rater reliability and concurrent validity of the AACPDM study design and quality rating system for conducting systematic	Refers to a tool to assess quality of published systematic reviews

Reference	Reason for exclusion
reviews (group design). Dev Med Child Neurol 2011;53:74.	

Table S3. General characteristics of included tools

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
Balshem 2013[1]	AHRQ outcome and analysis reporting bias framework	Domain-based	Reporting bias only	Bias due to selective non-reporting and bias in selection of the reported result	Randomized trials	Specific outcome/ result in a study	Expert consensus (via email)	Brief annotation per item/response option	No
Berkman 2013[2]	AHRQ tool for evaluating the risk of reporting bias	Domain-based	Reporting bias only	Bias due to selective publication and bias due to selective non-reporting	Systematic reviews	Specific synthesis of studies	Not stated	Brief annotation per item/response option	No
Downes 2016[3]	AXIS tool (Appraisal tool for Cross-Sectional Studies)	Checklist	Multiple sources of bias	Bias due to selective non-reporting	Cross-sectional studies	Whole study	Literature review, piloting, Delphi study	None	No
Downs 1998[4]	Downs-Black tool	Scale	Multiple sources of bias	Bias in selection of the	Randomized trials and non-	Whole study	Literature review, piloting,	Brief annotation per item/response	Yes

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
				reported result	randomized studies of interventions		psychometric testing	option	
Guyatt 2011[5-9]	GRADE	Domain-based	Multiple sources of bias	Bias due to selective publication and bias due to selective non-reporting	Systematic reviews	Specific synthesis of studies	Literature review, expert consensus (face-to-face and email), user testing	Detailed guidance manual	Yes
Hayden 2013[10]	QUIPS (Quality In Prognosis Studies) tool	Domain-based	Multiple sources of bias	Bias due to selective non-reporting	Prognosis studies	Whole study	Modified Delphi approach, nominal group technique at facilitated discussion workshop; piloting	Brief annotation per item/response option	Yes
Higgins 2008[11-13]	Cochrane risk of bias tool for randomized trials	Domain-based	Multiple sources of bias	Bias due to selective non-reporting and bias in selection of the reported	Randomized trials	Whole study	Literature review, informal consensus at facilitated meeting, piloting, focus groups and	Detailed guidance manual	Yes

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
				result			surveys, followed by consensus meeting		
Higgins 2016[14 15]	RoB 2.0 (revised tool for assessing risk of bias in randomized trials)	Domain-based	Multiple sources of bias	Bias in selection of the reported result	Randomized trials	Specific outcome/ result in a study	Literature review, informal consensus at facilitated meeting, piloting	Detailed guidance manual	No
Hoojimans 2014[16]	SYRCLE's RoB tool (SYstematic Review Centre for Laboratory animal Experimentation)	Domain-based	Multiple sources of bias	Bias due to selective non-reporting and bias in selection of the reported result	Animal studies	Whole study	Adaptation of existing tool, literature review	Brief annotation per item/response option	No
Kim 2013[17]	RoBANS (Risk of Bias Assessment Tool for Nonrandomized Studies)	Domain-based	Multiple sources of bias	Bias due to selective non-reporting and bias in selection of the	Non-randomized studies of interventions	Whole study	Literature review, psychometric testing	Brief annotation per item/response option	Yes

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
				reported result					
Kirkham 2010[18 19]	ORBIT-I (Outcome Reporting Bias In Trials) classification system for benefit outcomes	Domain-based	Reporting bias only	Bias due to selective non-reporting	Randomized trials	Specific outcome/ result in a study	Iteratively developed as part of a methodological study	Worked example for each response option	Yes
Meader 2014[20 21]	SAQAT (Semi-Automated Quality Assessment Tool)	Domain-based	Multiple sources of bias	Bias due to selective publication and bias due to selective non-reporting	Systematic reviews	Specific synthesis of studies	Development of logic model based on GRADE articles and piloting	None	Yes
Reid 2015[22]	Selective reporting bias algorithm	Domain-based	Reporting bias only	Bias due to selective non-reporting and bias in selection of the reported	Randomized trials	Whole study	Not stated	Brief annotation per item/response option	No

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
				result					
Saini 2014[23]	ORBIT-II (Outcome Reporting Bias In Trials) classification system for harm outcomes	Domain-based	Reporting bias only	Bias due to selective non-reporting	Randomized trials and non-randomized studies of interventions	Specific outcome/ result in a study	Iteratively developed as part of a methodological study	Worked example for each response option	No
Salanti 2014[24 25]	Framework for evaluating the quality of evidence from a network meta-analysis	Domain-based	Multiple sources of bias	Bias due to selective publication and bias due to selective non-reporting	Network meta-analyses	Specific synthesis of studies	Adaptation of existing tool	Detailed annotation per item/response option	No
Sterne 2016[26]	ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool	Domain-based	Multiple sources of bias	Bias in selection of the reported result	Non-randomized studies of interventions	Specific outcome/ result in a study	Expert consensus meetings (face-to-face), piloting	Detailed guidance manual	Yes

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
Viswanathan 2012[27]	RTI Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures	Domain-based	Multiple sources of bias	Bias due to selective non-reporting	Non-randomized studies of interventions or exposures	Whole study	Literature review, expert consensus (via email), cognitive testing, psychometric testing	Brief annotation per item/response option	No
Viswanathan 2013[28]	RTI Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures	Domain-based	Multiple sources of bias	Bias due to selective non-reporting	Non-randomized studies of interventions or exposures	Whole study	Literature review, expert consensus (via email)	Brief annotation per item/response option	No

References

1. Balshem H, Stevens A, Ansari M, et al. Finding grey literature evidence and assessing for outcome and analysis reporting biases when comparing medical interventions: AHRQ and the Effective Health Care Program. (Prepared by the Oregon Health and Science University and the University of Ottawa Evidence-based Practice Centers under Contract Nos. 290-2007-10057-I and 290-2007-10059-I.) AHRQ Publication No. 13(14)-EHC096-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

2. Berkman ND, Lohr KN, Ansari M, et al. Chapter 15 Appendix A: A Tool for Evaluating the Risk of Reporting Bias (in Chapter 15: Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update). Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I). AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm

3. Downes MJ, Brennan ML, Williams HC, et al. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ open* 2016;6:e011458.

4. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.

5. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.

6. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64(12):1277-82.

7. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol* 2011;64(4):407-15.

8. Schünemann H, Brożek J, Guyatt G, et al. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. [Updated October 2013]. Available from http://gdt.guidelinedevelopment.org/app/handbook/handbook.html.

9. Santesso N, Carrasco-Labra A, Langendam M, et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. *J Clin Epidemiol* 2016

10. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.

11. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons 2008:187-241.

12. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org/.

13. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

14. Higgins JPT, Savović J, Page MJ, et al. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0), Version 20 October 2016. Available from http://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/rob2-0/ [accessed 19 September 2017].

15. Higgins JPT, Sterne JAC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials. *Cochrane Methods Cochrane Database of Systematic Reviews* 2016;10(Suppl 1):29-31.

16. Hooijmans CR, Rovers MM, de Vries RBM, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.

17. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013;66(4):408-14.

18. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.

19. Dwan K, Gamble C, Kolamunnage-Dona R, et al. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials* 2010;11:52.

20. Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic reviews* 2014;3(1):82.

21. Stewart GB, Higgins JP, Schunemann H, et al. The use of Bayesian networks to assess the quality of evidence from research synthesis: 1. *PLoS One* 2015;10(3):e0114497.

22. Reid EK, Tejani AM, Huan LN, et al. Managing the incidence of selective reporting bias: a survey of Cochrane review groups. *Systematic reviews* 2015;4:85.

23. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ* 2014;349:g6501.

24. Salanti G, Giovane CD, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9(7):e99682.

25. Higgins JP, Del Giovane C, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *Value Health* 2014;17(7):A324.

26. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.

27. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* 2012;65(2):163-78.

28. Viswanathan M, Berkman ND, Dryden DM, et al. AHRQ Methods for Effective Health Care. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or

Exposures: Further Development of the RTI Item Bank. Rockville (MD): Agency for Healthcare Research and Quality (US) 2013.

Table S4. Items and response options relating to risk of reporting biases

Article ID	Tool	Items	Response options
Balshem 2013[1]	AHRQ outcome and analysis reporting bias framework	1. Across all study source documents, what is the risk of ORB/ARB? Compare published report(s) against (1) study protocol (if not retrieved in literature search), (2) trial registry entry/regulatory documents/industry documents, (3) other sources if applicable. 2. If ORB risk unclear: Given the study objectives, duration, and other investigated outcomes, could the study have also likely measured the outcome of interest but not reported it?	**Outcome reporting bias risk positive (ORB risk +)**: If reviewers determine that an outcome X was planned but the results were not reported, or were only partially reported in study documents, then the study is at risk of reporting bias for that outcome ("ORB risk +"). Also, if reviewers determine that an outcome X was not planned but the results were reported, then the study is at risk of reporting bias for that outcome ("ORB risk +"). Also, for studies for which the risk of reporting bias cannot be ruled out, reviewers should ask the question: "Given the study objectives, duration, and other investigated outcomes, could the study have also likely measured the outcome of interest but not reported it?" When the answer is "yes" (e.g., another reported outcome in the study leads the reviewer to believe that outcome X would have been collected), then the study should be rated "ORB risk +" for that outcome. **Outcome reporting bias risk negative (ORB risk -)**: When it is clear to the reviewers that outcome X was planned (e.g. from protocol, regulatory submissions, etc.), complete outcome data are available from at least one study document (published or otherwise), and the outcome was appropriately analyzed as planned, then the study is not at risk for reporting bias for this outcome. Also, for studies for which the risk of reporting bias cannot be ruled out, reviewers should ask the question: "Given the study objectives, duration, and other investigated outcomes, could the study have also

Article ID	Tool	Items	Response options
			likely measured the outcome of interest but not reported it?” If the answer is “no” the study should be rated as “ORB risk–”. **Outcome reporting bias risk unclear (ORB risk unclear)**: If the reviewers are unable to determine whether an outcome X was planned, but data are reported completely or partially, then the study risk of outcome and analysis reporting bias may be categorized as “unclear”. This would also apply to a study that did not report any outcome of review interest across all source documents but was eligible on population, intervention, comparator, and other criteria. Also, for studies for which the risk of reporting bias cannot be ruled out, reviewers should ask the question: “Given the study objectives, duration, and other investigated outcomes, could the study have also likely measured the outcome of interest but not reported it?” If it still remains unclear whether the outcome of interest may have been assessed, the study should be categorized as “ORB risk unclear.” **Analysis reporting bias risk positive (ARB risk +)**: When reported results are based on a different analysis, effect measure, cut-off, etc. than what was prespecified, then the study is at risk of analysis reporting bias for that outcome (“ARB risk +”). A study is also at risk of analysis reporting (“ARB risk +”) because there is no way to know whether the reported analysis was planned or post hoc. **Analysis reporting bias risk negative (ARB risk -)**: When it is clear to the reviewers that outcome X was planned (e.g. from protocol, regulatory submissions, etc.),

Article ID	Tool	Items	Response options
			complete outcome data are available from at least one study document (published or otherwise), and the outcome was appropriately analyzed as planned, then the study is not at risk for reporting bias for this outcome **Analysis reporting bias risk unclear (ARB risk unclear)**: If the reviewers are unable to determine whether an outcome X was planned, but data are reported completely or partially, then the study risk of outcome and analysis reporting bias may be categorized as "unclear". This would also apply to a study that did not report any outcome of review interest across all source documents but was eligible on population, intervention, comparator, and other criteria.
Berkman 2013[2]	AHRQ tool for evaluating the risk of reporting bias	1. Are all the following criteria met: ≥10 studies contributing data for an outcome, studies of unequal sizes, no substantial clinical and methodological differences between smaller and larger studies, and quantitative results accompanied with measures of dispersion? 2. If yes, do smaller studies tend to demonstrate more favorable results? (visual assessment) 3. If yes, what is the result of a test for funnel plot asymmetry? 4. If test is positive, would a clinical decision differ for estimates from a fixed effects versus random effect model because the findings from a fixed effect model are closer to the null? 5. If no to the first question, is there an explanation for substantial heterogeneity? 6. If no to any of Q1-5, what is the estimated N of studies that are affected by SOR, SAR,	**Suspected risk of reporting bias**: Testing for funnel plot asymmetry demonstrates a substantial likelihood of bias, and/or a qualitative assessment suggests the likelihood of missing studies, analyses, or outcomes data that may alter the conclusions from the reported evidence. **Undetected risk of reporting bias**: All alternative scenarios.

Article ID	Tool	Items	Response options
		nonpublication, or nonaccessibility? 7. If no to any of Q1-5, what is the total sample size of evidence affected by reporting bias (when known)? 8. If no to any of Q1-5, what is the total N of studies in evidence base? 9. If no to any of Q1-5, what is the total N of participants in evidence base? 10. If no to any of Q1-5, what is the consistency of effect estimates across contributing studies? 11. If no to any of Q1-5, what are the study limitations for the evidence base? 12. If no to any of Q1-5, what is the comprehensiveness of study retrieval and identification?	
Downes 2016[3]	AXIS tool (Appraisal tool for Cross-Sectional Studies)	1. Were the results for the analyses described in the methods, presented?	**Yes**: Not stated **No**: Not stated **Do not know/comment**: Not stated
Downs 1998[4]	Downs-Black tool	1. If any of the results of the study were based on "data dredging", was this made clear?	**Yes**: Any analyses that had not been planned at the outset of the study were clearly indicated. Also, no retrospective unplanned subgroup analyses were reported. **No**: Any analyses that had not been planned at the outset of the study were not clearly indicated. **Unable to determine**: Not stated
Guyatt 2011[5-9]	GRADE	1. Study limitations (including selective outcome reporting) 2. Publication bias	**Study limitations domain – No serious limitations, do not downgrade**: Most information is from studies at low risk of bias (i.e. those with low risk of bias for all key criteria, including lack of allocation concealment, lack of blinding, incomplete accounting of patients and outcome

Article ID	Tool	Items	Response options
			events, selective outcome reporting bias, other limitations [stopping early for benefit, use of unvalidated outcome measures, carryover effects in crossover trial, recruitment bias in cluster-randomized trial]) **Study limitations domain – Serious limitations, rate down one level (i.e., from high to moderate quality)**: Most information is from studies at moderate risk of bias **Study limitations domain – Very serious limitations, rate down two levels (i.e., from high to low quality or moderate to very low)**: Most information is from studies at high risk of bias. Selective reporting is present if authors acknowledge prespecified outcomes that they fail to report or report outcomes incompletely such that they cannot be included in a metaanalysis. One should suspect reporting bias if the study report fails to include results for a key outcome that one would expect to see in such a study or if composite outcomes are presented without the individual component outcomes. **Publication bias domain – Undetected**: None of the criteria for "strongly suspected" are met **Publication bias domain – Strongly suspected**: "In general, review authors and guideline developers should consider rating down for likelihood of publication bias when the evidence consists of a number of small studies. The inclination to rate down for publication bias should increase if most of those small studies are industry sponsored or likely to be industry sponsored (or if the investigators share another conflict of interest)...Another criterion for publication bias is the pattern of study results. Suspicion may increase if visual inspection

Article ID	Tool	Items	Response options
			demonstrates an asymmetrical rather than a symmetrical funnel plot or if statistical tests of asymmetry are positive. Although funnel plots may be helpful, review authors and guideline developers should bear in mind that visual assessment of funnel plots is distressingly prone to error. Enhancements of funnel plots may (or may not) help to improve reproducibility and validity associated with their use...Furthermore, systematic review and guideline authors should bear in mind that even if they find convincing evidence of asymmetry, publication bias is not the only explanation. For instance, if smaller studies suffer from greater study limitations, they may yield biased overestimates of effects. Another explanation would be that, because of a more restrictive (and thus responsive) population, or a more careful administration of the intervention, the effect may actually be larger in the small studies...More compelling than any of these theoretical exercises is authors' success in obtaining the results of some unpublished studies and demonstrating that the published and unpublished data show different results. In these circumstances, the possibility of publication bias looms large. The risk of publication bias is probably larger for observational studies than for RCTs, particularly small observational studies and studies conducted on data collected automatically (e.g. in the electronic medical record or in a diabetes registry) or data collected for a previous study. In these instances, it is difficult for the reviewer to know if the observational studies that appear in the literature represent all or a fraction of the studies conducted, and whether the analyses in them represent all or a fraction of those

Article ID	Tool	Items	Response options
			conducted. In these instances, reviewers may consider the risk of publication bias as substantial" [6]. "Guideline panels and authors of systematic reviews should consider the extent to which they are uncertain about the magnitude of the effect due to selective publication of studies and they may downgrade the quality of evidence by one level. Consider: study design (experimental vs. observational); study size (small studies vs. large studies); lag bias (early publication of positive results); search strategy (was it comprehensive?); asymmetry in funnel plot" [8]. "Relevant content: whether publication bias is undetected or suspected; interpretation of funnel plot; comprehensiveness of the search strategies and methods to identify all available evidence; presence of small (often positive) studies with for profit interest...Indicate the reason publication bias is detected (e.g. asymmetrical funnel plot, small studies with positive results, suspected selective availability of data from published, or unpublished studies)" [9].
Hayden 2013[10]	QUIPS (Quality In Prognosis Studies) tool	1. Statistical analysis and reporting (the statistical analysis is appropriate and all primary outcomes are reported). Prompting items include (a) Sufficient presentation of data to assess the adequecy of the analytic strategy; (b) Strategy for model building is appropriate and is based on a conceptual framework or model; (c) The selected statistical model is adequate for the design of the study; (d) There is no selective reporting of results.	**Low risk of bias**: The reported results are unlikely to be spurious or biased related to analysis or reporting **Moderate risk of bias**: The reported results may be spurious or biased related to analysis or reporting **High risk of bias**: The reported results are very likely to be spurious or biased related to analysis or reporting
Higgins	Cochrane risk of bias tool for	1. Are reports of the study free of suggestion of selective outcome reporting? (2008 version);	**Low risk of bias**: Any of the following – The study protocol is available and all of the study's pre-specified

Article ID	Tool	Items	Response options
2008[11-13]	randomized trials	Reporting bias due to selective outcome reporting (2011 version)	(primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way; The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon). **High risk of bias**: Any one of the following – Not all of the study's pre-specified primary outcomes have been reported; One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not prespecified; One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; The study report fails to include results for a key outcome that would be expected to have been reported for such a study. **Unclear risk of bias**: Insufficient information to permit judgement of 'Low risk' or 'High risk'. It is likely that the majority of studies will fall into this category.
Higgins 2016[14 15]	RoB 2.0	1. Are the reported outcome data likely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain, or from multiple analyses of the data?	**Low risk of bias**: Reported outcome data are unlikely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain, and reported outcome data are unlikely to have been selected, on the basis of the results, from multiple analyses of the data. **High risk of bias**: Reported outcome data are likely to have been selected, on the basis of the results, from

Article ID	Tool	Items	Response options
			multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain, or from multiple analyses of the data (or both). **Some concerns**: There is insufficient information available to exclude the possibility that reported outcome data were selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain, or from multiple analyses of the data.
Hoojimans 2014[16]	SYRCLE's RoB tool (SYstematic Review Centre for Laboratory animal Experimentation)	1. Are reports of the study free of selective outcome reporting? Includes two signalling questions: Was the study protocol available and were all of the study's pre-specified primary and secondary outcomes reported in the current manuscript?; Was the study protocol not available, but was it clear that the published report included all expected outcomes (i.e. comparing methods and results section)?	**Low risk of bias**: Not stated, but assume same criteria as Cochrane risk of bias tool for randomized trials [13]. **High risk of bias**: Not all of the study's pre-specified primary outcomes have been reported; One or more primary outcomes have been reported using measurements, analysis methods or data subsets (e.g. subscales) that were not pre-specified in the protocol; One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting has been provided, such as an unexpected adverse effect); The study report fails to include results for a key outcome that would be expected to have been reported for such a study. **Unclear risk of bias**: Not stated, but assume same criteria as Cochrane risk of bias tool for randomized trials [13].
Kim 2013[17]	RoBANS (Risk of Bias Assessment Tool for Nonrandomized	1. Reporting biases caused by the selective reporting of outcomes	**Low risk of bias**: Any one of the following conditions – The experimental protocol is available, and the pre-defined primary/secondary outcomes were described as planned; All of the expected outcomes were included in

Article ID	Tool	Items	Response options
	Studies)		the study descriptions (even in the absence of the experimental protocols). **High risk of bias**: Any one of the following conditions – The pre-defined primary outcomes were not fully reported; The outcomes were not reported in accordance with the previously defined standards; Primary outcomes that were not pre-specified in the study existed (except for outcomes with clear explanations, such as unexpected adverse effects); The existence of incomplete reporting regarding the primary outcome of interest; The absence of reports on important outcomes that would be expected to be reported for studies in related fields. **Unclear risk of bias**: It is uncertain whether the selective outcome reporting resulted in a 'high risk' or a 'low risk' of bias.
Kirkham 2010[18 19]	ORBIT-I (Outcome Reporting Bias In Trials) classification system for benefit outcomes	1. The Outcome Reporting Bias In Trials (ORBIT) study classification system for missing or incomplete outcome reporting in reports of randomised trials	**Low risk of bias**: A "low risk" classification was awarded when it was suspected, but not actually known, that the outcome was either not measured, measured but not analysed, or measured and analysed but either partially reported or not reported for a reason unrelated to the results obtained. Specific examples include: (C) Trial report states that outcome was analysed but insufficient data were presented for the trial to be included in meta-analysis or to be considered to be fully tabulated; (F) Clear that outcome was measured but not necessarily analysed, and judgment says unlikely to have been analysed but not reported because of non-significant results; (H) Not mentioned but clinical judgment says outcome unlikely to have been measured at all.

Article ID	Tool	Items	Response options
			High risk of bias: A "high risk" classification was awarded when it was either known or suspected that the results were partially or not reported because the treatment comparison was statistically non-significant (P>0.05). Specific examples include: (A) Trial report states that outcome was analysed but only reports that result was not significant (typically stating P>0.05); (D) Trial report states that outcome was analysed but no results reported; (E) Clear that outcome was measured but not necessarily analysed, and judgment says likely to have been analysed but not reported because of non-significant results; (G) Not mentioned but clinical judgment says outcome likely to have been measured and analysed but not reported on the basis of non-significant results. **No risk of bias**: A "no risk" classification was reserved for cases where it was known that the outcome was not measured, known that it was measured but not analysed, or known that it was measured and analysed but the reason for partial or no reporting was not because the results were statistically non-significant. Specific examples include: (B) Trial report states that outcome was analysed but only reports that result was significant (typically stating P<0.05); (I) Clear that outcome was not measured.
Meader 2014[20 21]	SAQAT (Semi-Automated Quality Assessment Tool)	Study limitations domain 1. Were data reported consistently for the outcome of interest (i.e. no potential selective reporting)? Publication bias domain	**Study limitations domain – No serious limitations**: No problem for any source of risk of bias. **Study limitations domain – Serious limitations**: Selection bias results in serious limitations, or very serious limitations if combined with a problem from any

Article ID	Tool	Items	Response options
		1. Did the authors conduct a comprehensive search? 2. Did the authors search for grey literature? 3. Authors did not apply restrictions to study selection on the basis of language? 4. There was no industry influence on studies included in the review? 5. There was no evidence of funnel plot asymmetry? 6. There was no discrepancy in findings between published and unpublished trials?	alternative source; two problems from other sources (e.g. detection bias, attrition bias) result in serious limitations. **Study limitations domain – Very serious limitations**: Selection bias results in serious limitations, or very serious limitations if combined with a problem from any alternative source; three problems result in very serious limitations **Publication bias domain – Strongly suspected**: High probability of publication bias. Responses to each item are entered into a Bayesian network to ascertain the probabilities of each GRADE domain. Publication bias is determined by a combination of discrepancy between published and unpublished studies (yes/no), amount of statistical information (high/intermediate/low), industry influence (yes/no) and search integrity (high/low), with the former carrying greatest weight. That is, the probability of publication bias is always considered high when there is a discrepancy between published and unpublished studies (regardless of responses to other items). **Publication bias domain – Undetected**: Low probability of publication bias (as determined by the Bayesian network described above.
Reid 2015[22]	Selective reporting bias algorithm	1. Protocol available? 2. Trial registration? 3. Outcomes described? 4. Response from contact with study authors? 5. Outcomes match?	**High risk of bias**: Outcomes are described in the protocol or trial registry or by the review authors when contacted, and they do not match the outcomes reported. **Low risk of bias**: Outcomes are described in the protocol or trial registry or by the review authors when contacted,

Article ID	Tool	Items	Response options
			and they do match the outcomes reported. **Unclear risk of bias**: Outcomes are not described in the protocol or trial registry, or a protocol or trial registry are not available and no response is received from review authors when contacted.
Saini 2014[23]	ORBIT-II (Outcome Reporting Bias In Trials) classification system for harm outcomes	1. ORBIT-II classification system	**Low risk of bias**: Specific examples include: (P3) Explicit specific harm measured and compared across treatment groups, although insufficient reporting for meta-analysis or full tabulation; (T1) Clinical judgement says specific harm likely measured but no events, because specific harm not mentioned but all other specific harms fully reported; (T2) Clinical judgement says specific harm likely measured but no events, because there was no description of specific harms; (U) Specific harm outcome not explicitly mentioned, clinical judgment says unlikely measured (no harms mentioned or reported). **High risk of bias**: In the context of harm outcomes, we awarded classifications for "high risk" outcome reporting bias when the specific harm had been measured but the data were presented or suppressed in a way that would mask the harm profile of particular interventions (including providing detail on the seriousness of the harms)—that is, P1, P2, R, and S classifications. Specific examples include: (P1) States outcome analysed but reported only that P>0.05; (P2) States outcome analysed but reported only that P<0.05; (R1) Clear that outcome was measured but no results reported; (R2) Result reported globally across all groups; (R3) Result reported from some groups only; (S1) Clinical judgment says specific harm outcome likely measured and likely

Article ID	Tool	Items	Response options
			compared across treatment groups, but only pooled adverse events reported (could include specific harm outcome); (S2) Clinical judgment says specific harm outcome likely measured and likely compared across treatment groups, but no harms mentioned or reported. **No risk of bias**: Specific examples include: (Q) Clear that explicit specific harm outcome was measured and clear outcome was not compared; (V) Report clearly specifies that data on specific harm of interest was not measured.
Salanti 2014[24 25]	Framework for evaluating the quality of evidence from a network meta-analysis	1. Study limitations (including selective outcome reporting) evaluated in a specific pairwise effect estimated in network meta-analysis: Determine which direct comparisons contribute to estimation of the NMA treatment effect and integrate risk of bias assessments from these into a single judgment. 2. Publication bias evaluated in a specific pairwise effect estimated in network meta-analysis: Non-statistical consideration of likelihood of non-publication of evidence that would inform the pairwise comparison. Plot pairwise estimates on contour-enhanced funnel plot. 3. Study limitations (including selective outcome reporting) evaluated in treatment ranking estimated in network meta-analysis: Integrate risk of bias assessments from each direct comparison to formulate a single overall confidence rating for treatment rankings. 4. Publication bias evaluated in treatment ranking estimated in network meta-analysis: Non-statistical consideration of likelihood of non-publication for each pairwise comparison. If appropriate, plot NMA	**Study limitations domain – No serious limitations, do not downgrade**: Use standard GRADE considerations to inform judgment [7]. **Study limitations domain – Serious limitations, rate down one level (i.e., from high to moderate quality)**: Use standard GRADE considerations to inform judgment [7]. **Study limitations domain – Very serious limitations, rate down two levels (i.e., from high to low quality or moderate to very low)**: Use standard GRADE considerations to inform judgment [7]. **Publication bias domain (evaluated in a specific pairwise effect estimated in network meta-analysis) – Undetected**: Use standard GRADE to inform judgment [6]. **Publication bias domain (evaluated in a specific pairwise effect estimated in network meta-analysis) – Strongly suspected**: "Even after a meticulous search for studies, publication bias can occur and usually it tends to lead to overestimation of an active treatment's effect compared with placebo or other reference treatment.

Article ID	Tool	Items	Response options
		estimates on a comparison adjusted funnel plot and assess asymmetry.	Several approaches have been proposed to generate assumptions about the presence of publication bias, including funnel plots, regression methods and selection models, but each has limitations and their appropriateness is often debated. Making judgements about the presence of publication bias in a network meta-analysis is usually difficult. We suggest that for each observed pairwise comparison, judgements about the presence of publication bias are made using standard GRADE. We recommend that the primary considerations are non-statistical (by considering how likely it is that studies may have been performed but not published) and we advocate the use of contour-enhanced funnel plots, which may help in identifying publication bias as a likely explanation of funnel plot asymmetry. Then, judgements about the direct effects can be summarized to infer about the network estimates by taking into account the contributions of each direct piece of evidence" [24]. **Publication bias domain (evaluated in treatment ranking estimated in network meta-analysis) – Undetected**: Use standard GRADE to inform judgment [6]. **Publication bias domain (evaluated in treatment ranking estimated in network meta-analysis) – Strongly suspected**: “Judgments about the potential impact of publication bias in the ranking of the treatments require, as before, consideration of the comprehensiveness of the search for studies and the likelihood that studies may have been conducted and not published. A statistical approach to detecting bias is offered in certain situations by the comparison-adjusted funnel plot for a network of

Article ID	Tool	Items	Response options
			treatments. In such a plot, the vertical axis represents the inverted standard error of the effect sizes as in a standard funnel plot. However, the horizontal axis represents an adjusted effect size, presenting the difference between each observed effect size and the mean effect size for the specific comparison being made. The use of such a plot is informative only when the comparisons can confidently be ordered in a meaningful way; for example, if all comparisons are of active treatment versus placebo, or all are of a new versus an old drug. Examination of any asymmetry in the plot can help to infer about the possible presence of an association between study size and study effect. Asymmetry does not provide evidence of publication bias, however, since associations between effect size and study size can be due to study limitations or genuine heterogeneity of effects” [24].
Sterne 2016[26]	ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool	1. Is the reported effect estimate likely to be selected, on the basis of the results, from multiple outcome measurements within the outcome domain, multiple analyses of the intervention-outcome relationship, or different subgroups?	**Low risk of bias**: There is clear evidence (usually through examination of a pre-registered protocol or statistical analysis plan) that all reported results correspond to all intended outcomes, analyses and subcohorts. **Moderate risk of bias**: (i) The outcome measurements and analyses are consistent with an a priori plan; or are clearly defined and both internally and externally consistent; and (ii) There is no indication of selection of the reported analysis from among multiple analyses; and (iii) There is no indication of selection of the cohort or subgroups for analysis and reporting on the basis of the results. **Serious risk of bias**: (i) Outcomes are defined in different

Article ID	Tool	Items	Response options
			ways in the methods and results sections, or in different publications of the study; or (ii) There is a high risk of selective reporting from among multiple analyses; or (iii) The cohort or subgroup is selected from a larger study for analysis and appears to be reported on the basis of the results. **Critical risk of bias**: (i) There is evidence or strong suspicion of selective reporting of results; and (ii) The unreported results are likely to be substantially different from the reported results. **No information**: There is too little information to make a judgement (for example, if only an abstract is available for the study).
Viswanathan 2012[27]	RTI Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures	1. Are any important primary outcomes missing from the results? 2. Are any important harms or adverse events that may be a consequence of the intervention/exposure missing from the results?	**Yes (for item on primary outcome)**: No specific criteria stated. Only guidance is "Identify all primary outcomes, including timing of measurement, that one would expect to be reported in the study" **No (for item on primary outcome)**: No specific criteria stated. **Cannot determine (for item on primary outcome)**: No specific criteria stated. **Yes (for item on harm outcome)**: No specific criteria stated. Only guidance is "Identify all important harms, including timing of measurement, that one would expect be reported in the study. Drop if not relevant to body of literature." **Partially (for item on harm outcome)**: No specific criteria stated.

Article ID	Tool	Items	Response options
			No (for item on harm outcome): No specific criteria stated. **Assessment of harms not applicable to this study (for item on harm outcome)**: No specific criteria stated.
Viswanathan 2013[28]	RTI Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures	1. Are any important primary outcomes missing from the results? 2. Are any important harms or adverse events that may be a consequence of the intervention/exposure missing from the results?	**Yes, important outcome(s) missing (for item on primary outcome)**: No specific criteria stated. Only guidance is "Identify all primary outcomes that one would expect to be reported in the study, including timing of measurement." **No important outcome (s) missing (for item on primary outcome)**: No specific criteria stated. **Cannot determine (for item on primary outcome)**: No specific criteria stated. **Yes, important outcomes missing (for item on harm outcome)**: No specific criteria stated. Only guidance is "Identify all important harms that one would expect be reported in the study, including timing of measurement. Drop if not relevant to body of literature." **No important outcomes missing (for item on harm outcome)**: No specific criteria stated. **Assessment of harms not applicable to this study (for item on harm outcome)**: No specific criteria stated.

References

1. Balshem H, Stevens A, Ansari M, et al. Finding grey literature evidence and assessing for outcome and analysis reporting biases when comparing medical interventions: AHRQ and the Effective Health Care Program. (Prepared by the Oregon Health and Science University and the University of Ottawa Evidence-based Practice Centers under Contract Nos. 290-2007-10057-I and 290-2007-10059-I.) AHRQ Publication No. 13(14)-EHC096-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

2. Berkman ND, Lohr KN, Ansari M, et al. Chapter 15 Appendix A: A Tool for Evaluating the Risk of Reporting Bias (in Chapter 15: Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update). Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I). AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm

3. Downes MJ, Brennan ML, Williams HC, et al. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ open* 2016;6:e011458.

4. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.

5. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.

6. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64(12):1277-82.

7. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol* 2011;64(4):407-15.

8. Schünemann H, Brożek J, Guyatt G, et al. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. [Updated October 2013]. Available from http://gdt.guidelinedevelopment.org/app/handbook/handbook.html.

9. Santesso N, Carrasco-Labra A, Langendam M, et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. *J Clin Epidemiol* 2016

10. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.

11. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons 2008:187-241.

12. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org/.

13. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

14. Higgins JPT, Savović J, Page MJ, et al. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0), Version 20 October 2016. Available from http://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/rob2-0/ [accessed 19 September 2017].

15. Higgins JPT, Sterne JAC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials. *Cochrane Methods Cochrane Database of Systematic Reviews* 2016;10(Suppl 1):29-31.

16. Hooijmans CR, Rovers MM, de Vries RBM, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.

17. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013;66(4):408-14.

18. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.

19. Dwan K, Gamble C, Kolamunnage-Dona R, et al. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials* 2010;11:52.

20. Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic reviews* 2014;3(1):82.

21. Stewart GB, Higgins JP, Schunemann H, et al. The use of Bayesian networks to assess the quality of evidence from research synthesis: 1. *PLoS One* 2015;10(3):e0114497.

22. Reid EK, Tejani AM, Huan LN, et al. Managing the incidence of selective reporting bias: a survey of Cochrane review groups. *Systematic reviews* 2015;4:85.

23. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ* 2014;349:g6501.

24. Salanti G, Giovane CD, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9(7):e99682.

25. Higgins JP, Del Giovane C, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *Value Health* 2014;17(7):A324.

26. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.

27. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* 2012;65(2):163-78.

28. Viswanathan M, Berkman ND, Dryden DM, et al. AHRQ Methods for Effective Health Care. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or

Exposures: Further Development of the RTI Item Bank. Rockville (MD): Agency for Healthcare Research and Quality (US) 2013.

Table S5. General characteristics of studies evaluating the measurement properties of tools for assessing risk of reporting biases

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Armijo-Olivo 2012[1]	Cochrane risk of bias tool for randomized trials (2008 version)	None	20 trials included in a SR exploring knowledge transfer interventions for cancer pain management.	Cancer pain	None	20	NA	Range 1987-2007	2
Armijo-Olivo 2014[2]	Cochrane risk of bias tool for randomized trials (2011 version)	Inter-rater reliability	Trials of physical therapy interventions included in meta-analyses of a continuous outcome.	Physical therapy for musculoskeletal, cardiorespiratory, neurological or gynaecological conditions	None	109	NA	Not reported	2
Bilandzic 2016[3]	ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool	Inter-rater reliability	Studies included in two SRs of NRSI of the relationship between the use of TZDs and COX-2 inhibitors and major cardiovascular events.	Cardiovascular disease	None	37	NA	Range 2000-2010	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Downs 1998[4]	Downs-Black tool	None	10 randomised controlled trials and 10 non-randomised trials/prospective cohort studies randomly selected from studies identified during a SR of surgery for stress incontinence	Stress incontinence	None	20	NA	Not reported	2
Hartling 2009[5]	Cochrane risk of bias tool for randomized trials (2008 version)	Inter-rater reliability	A convenience sample of 163 randomized trial in child health, which were presented at the annual scientific meetings of the Society for Pediatric Research between 1992 and 1995.	Child health	None	163	NA	Not reported	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Hartling 2011[6]	Cochrane risk of bias tool for randomized trials (2008 version)	Inter-rater reliability	Trials included in a systematic review of long-acting beta agonists (LABA) combined with inhaled corticosteroids (ICS) for adults with persistent asthma.	Asthma	None	107	NA	Median 2004, IQR 2001-2006	2
Hartling 2012[7 8]	Cochrane risk of bias tool for randomized trials (2011 version)	Inter-rater reliability	A sample of 154 trial was randomly selected from among 616 trials published in December 2006 that were previously examined for quality of reporting.	Varied	None	154	NA	All 2006	2
Hayden 2013[9]	QUIPS (Quality In Prognosis Studies) tool	Inter-rater reliability	Studies included in a systematic review of troponin-based risk stratification of patients with	Pulmonary embolism	None	31	NA	Not reported	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
			acute non-massive pulmonary embolism.						
Hoojimans 2014[10]	SYRCLE's RoB tool (SYstematic Review Centre for Laboratory animal Experimentation)	Inter-rater reliability	1 systematic review including 32 papers (no other details provided).	Animal studies (not specified)	None	32	NA	Not reported	2
Jordan 2017[11]	Cochrane risk of bias tool for randomized trials (2011 version)	Inter-rater reliability	Any study that had been included more than once in SRs present on the Cochrane Database of Systematic Reviews in the area of subfertility.	Subfertility	None	28	NA	Not reported	2
Kim 2013[12]	RoBANS (Risk of Bias Assessment Tool for Nonrandomized Studies)	Inter-rater reliability	39 NRSs from four systematic reviews (one by the National Evidence-based Healthcare Collaborating Agency and three Cochrane reviews).	Depression, myocardial infarction, post-partum hemorrhage, chronic non-cancer pain	None	39	NA	Not reported	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Kumar 2016[13]	GRADE	None	10 key questions that were systematically reviewed for a clinical practice guideline for the use of prophylactic vs. therapeutic platelet transfusion in patients with thrombocytopenia.	Thrombocytopenia	10	None	All 2015	NA	18
Llewellyn 2015[14]	SAQAT (Semi-Automated Quality Assessment Tool)	Inter-rater reliability	29 meta-analyses from a purposive sample of SRs of RCTs from the Database of Systematic Reviews of Effects (DARE), and a purposive sample of 15 recent Cochrane reviews in mental health.	Varied	44	None	2006-2013	NA	2
Mustafa 2013[15]	GRADE	None	4 well-conducted and well-reported Cochrane reviews,	Alcohol dependence, asthma,	16	None	2004-2012	NA	4

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
			based on assessment using the AMSTAR tool.	cardiopulmonary bypass					
Norris 2012[16]	ORBIT-I (Outcome Reporting Bias In Trials) classification system for benefit outcomes	Inter-rater reliability; Time to complete assessments	Studies included in three AHRQ-funded comparative effectiveness reviews of randomised trials with drug-drug or drug-placebo comparisons, examining benefit outcomes.	Varied	None	40	NA	2005-2010	2
O'Connor 2015[17]	Downs-Black tool	None	20 studies included in an updated SR which examined the effects of an exercise intervention for chronic musculoskeletal pain.	Chronic musculoskeletal pain	None	20	NA	1997-2008	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Vale 2013[18]	Cochrane risk of bias tool for randomized trials (2011 version)	Agreement between assessments performed using published article only versus published article and data collected during the individual participant data process.	13 completed individual participant data meta-analyses of treatments for cancer. Trials had to be published either in full or as an abstract, and a copy of the trial protocol or forms detailing trial design completed by trialists (or both) had to be available.	Cancer pain	None	95	NA	Not reported	2

NA = Not applicable; SR = systematic review

References

1. Armijo-Olivo S, Stiles CR, Hagen NA, et al. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract* 2012;18(1):12-8.

2. Armijo-Olivo S, Ospina M, da Costa BR, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. *PLoS One* 2014;9(5):e96920.

3. Bilandzic A, Fitzpatrick T, Rosella L, et al. Risk of Bias in Systematic Reviews of Non-Randomized Studies of Adverse Cardiovascular Effects of Thiazolidinediones and Cyclooxygenase-2 Inhibitors: Application of a New Cochrane Risk of Bias Tool. *PLoS Med* 2016;13(4):e1001987.

4. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.

5. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.

6. Hartling L, Bond K, Vandermeer B, et al. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. *PLoS One* 2011;6(2):e17242.

7. Hartling L, Hamm M, Milne A, et al. AHRQ Methods for Effective Health Care. Validity and Inter-Rater Reliability Testing of Quality Assessment Instruments. Rockville (MD): Agency for Healthcare Research and Quality (US) 2012.

8. Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol* 2013;66(9):973-81.

9. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.

10. Hooijmans CR, Rovers MM, de Vries RBM, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.

11. Jordan VM, Lensen SF, Farquhar CM. There were large discrepancies in risk of bias tool judgments when a randomized controlled trial appeared in more than one systematic review. *J Clin Epidemiol* 2017;81:72-76.

12. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013;66(4):408-14.

13. Kumar A, Miladinovic B, Guyatt GH, et al. GRADE guidelines system is reproducible when instructions are clearly operationalized even among the guidelines panel members with limited experience with GRADE. *J Clin Epidemiol* 2016;75:115-8.

14. Llewellyn A, Whittington C, Stewart G, et al. The Use of Bayesian Networks to Assess the Quality of Evidence from Research Synthesis: 2. Inter-Rater Reliability and Comparison with Standard GRADE Assessment. *PLoS One* 2015;10(12):e0123511.

15. Mustafa RA, Santesso N, Brozek J, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol* 2013;66(7):736-42; quiz 42.e1-5.

16. Norris SL, Holmer HK, Ogden LA, et al. AHRQ Methods for Effective Health Care. Selective Outcome Reporting as a Source of Bias in Reviews of Comparative Effectiveness. Rockville (MD): Agency for Healthcare Research and Quality (US) 2012.

17. O'Connor SR, Tully MA, Ryan B, et al. Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. *BMC Res Notes* 2015;8:224.

18. Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. *BMJ* 2013;346:f1798.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Table S1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7-8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	NA
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	NA

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	NA
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8, Fig 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	12
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	NA
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Table S3 and S4
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	NA
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	13-22
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	23
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	23-24
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	26
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	26-27

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.

Page 2 of 2

BMJ Open

Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019703.R2
Article Type:	Research
Date Submitted by the Author:	10-Jan-2018
Complete List of Authors:	Page, Matthew; Monash University, School of Public Health and Preventive Medicine McKenzie, Joanne; Monash University, School of Public Health and Preventive Medicine Higgins, Julian; University of Bristol, Population Health Sciences, Bristol Medical School
<b>Primary Subject Heading</b>:	Research methods
Secondary Subject Heading:	Evidence based practice, Research methods
Keywords:	Publication Bias, Bias (Epidemiology), Review Literature As Topic, Checklist

Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review

Matthew J Page[1,2], Joanne E McKenzie[1], Julian PT Higgins[2]

1. School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

2. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

Correspondence to: Dr. Matthew Page, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne VIC 3004, Australia. Phone: +61 3 9903 0248. Email address: matthew.page@monash.edu

WORD COUNT: 4,139

ABSTRACT

BACKGROUND: Several scales, checklists and domain-based tools for assessing risk of reporting biases exist, but it is unclear how much they vary in content and guidance. We conducted a systematic review of the content and measurement properties of such tools.

METHODS: We searched for potentially relevant articles in Ovid MEDLINE, Ovid EMBASE, Ovid PsycINFO, and Google Scholar from inception to February 2017. One author screened all titles, abstracts and full text articles, and collected data on tool characteristics.

RESULTS: We identified 18 tools that include an assessment of the risk of reporting bias. Tools varied in regard to the type of reporting bias assessed (e.g. bias due to selective publication, bias due to selective non-reporting), and the level of assessment (e.g. for the study as a whole, a particular result within a study, or a particular synthesis of studies). Various criteria are used across tools to designate a synthesis as being at "high" risk of bias due to selective publication (e.g. evidence of funnel plot asymmetry, use of non-comprehensive searches). However, the relative weight assigned to each criterion in the overall judgement is unclear for most of these tools. Tools for assessing risk of bias due to selective non-reporting guide users to assess a study, or an outcome within a study, as "high" risk of bias if no results are reported for an outcome. However, assessing the corresponding risk of bias in a synthesis that is missing the non-reported outcomes is outside the scope of most of these tools. Inter-rater agreement estimates were available for five tools.

CONCLUSION: There are several limitations of existing tools for assessing risk of reporting biases, in terms of their scope, guidance for reaching risk of bias judgements, and measurement properties. Development and evaluation of a new, comprehensive tool, could help overcome present limitations.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- Tools for assessing risk of reporting biases, and studies evaluating their measurement properties, were identified by searching several relevant databases using a search string developed in conjunction with an information specialist.
- Detailed information on the content and measurement properties of existing tools was collected, providing readers with pertinent information to help decide which tools to use in evidence syntheses.
- Screening of articles and data collection were performed by one author only, so it is possible that some relevant articles were missed, or that errors in data collection were made.
- The search of grey literature was not comprehensive, so it is possible that there are other tools for assessing risk of reporting biases, and unpublished studies evaluating measurement properties, that were omitted from this review.

BACKGROUND

The credibility of evidence syntheses can be compromised by reporting biases, which arise when dissemination of research findings is influenced by the nature of the results[1]. For example, there may be bias due to selective publication, where a study is only published if the findings are considered interesting (also known as publication bias)[2]. In addition, bias due to selective non-reporting may occur, where findings (e.g. estimates of intervention efficacy or an association between exposure and outcome) that are statistically non-significant are not reported or are partially reported in a paper (e.g. stating only that "P>0.05")[3]. Alternatively, there may be bias in selection of the reported result, where authors perform multiple analyses for a particular outcome/association, yet only report the result which yielded the most favourable effect estimate[4]. Evidence from cohorts of clinical trials followed from inception suggest that biased dissemination is common. Specifically, on average, half of all trials are not published[1 5], trials with statistically significant results are twice as likely to be published[5], and a third of trials have outcomes that are omitted, added or modified between protocol and publication[6].

Audits of systematic review conduct suggest that most systematic reviewers do not assess risk of reporting biases[7-10]. For example, in a cross-sectional study of 300 systematic reviews indexed in MEDLINE® in February 2014[7], the risk of bias due to selective publication was not considered in 56% of reviews. A common reason for not doing so was that the small number of included studies, or inability to perform a meta-analysis, precluded the use of funnel plots. Only 19% of reviews included a search of a trial registry to identify completed but unpublished trials or pre-specified but non-reported outcomes, and only 7% included a search of another source of data disseminated outside of journal articles. The risk of bias due to selective non-reporting in the included studies was assessed in only 24% of reviews[7]. Another study showed that authors of Cochrane reviews routinely record whether any outcomes that were measured were not reported in the included trials, yet rarely consider if such non-reporting could have biased the results of a synthesis[11].

Previous researchers have summarised the characteristics of tools designed to assess various sources of bias in randomized trials[12-14], non-randomized studies of interventions (NRSI)[14 15], diagnostic test accuracy studies[16], and systematic reviews[14 17]. Others have summarised the performance of statistical methods developed to detect or adjust for reporting biases[18-20]. However, no prior review has focused specifically on tools (i.e. structured instruments such as scales, checklists, or domain-based tools) for assessing the risk of reporting biases. A particular challenge when assessing risk of reporting biases is that existing tools vary in their level of assessment. For example, tools for assessing risk of bias due to selective publication direct assessments at the level of the synthesis, whereas tools for assessing risk of bias due to selective non-reporting within studies can direct assessments at the level of the individual study, at the level of the synthesis, or at both levels. It is unclear how many tools are available to assess different types of reporting bias, and what level they direct assessments at. It is also unclear whether criteria for reaching risk of bias judgements are consistent across existing tools. Therefore, the aim of this research was to conduct a systematic review of the content and measurement properties of such tools.

METHODS

Protocol

Methods for this systematic review were pre-specified in a protocol, which was uploaded to the Open Science Framework in February 2017 (https://osf.io/9ea22/).

Eligibility criteria

Papers were included if the authors described a tool that was designed for use by individuals performing evidence syntheses to assess risk of reporting biases in the included studies or in their synthesis of studies. Tools could assess any type of reporting bias, including bias due to selective publication, bias due to selective non-reporting, or bias in selection of the reported result. Tools could assess the risk of reporting biases in any type of study (e.g. randomized trial of intervention, diagnostic test accuracy study, observational study estimating prevalence of an exposure), and in any type of result (e.g. estimate of intervention efficacy or harm, estimate of diagnostic accuracy, association between exposure and outcome). Eligible tools could take any form, including scales, checklists, and domain-based tools. To be considered a scale, each item had to have a numeric score attached to it, so that an overall summary score could be calculated[12]. To be considered a checklist, the tool had to include multiple questions, but the developers' intention was not to attach a numerical score to each response, or to calculate an overall score[13]. Domain-based tools were those that required users to judge risk of bias or quality within specific domains, and to record the information on which each judgement was based[21].

Tools with a broad scope, for example, to assess multiple sources of bias or the overall quality of the body of evidence, were eligible if one of the items covered risk of reporting bias. Multi-dimensional tools with a statistical component were also eligible (e.g. those that require users to respond to a set of questions about the comprehensiveness of the search, as well as to perform statistical tests for funnel plot asymmetry). In addition, any studies that evaluated the measurement properties of existing tools (e.g. construct validity, inter-rater agreement, time taken to complete assessments) were eligible for inclusion. Papers were eligible regardless of the date or format of publication, but were limited to those written in English.

The following were ineligible:

- articles or book chapters providing guidance on how to address reporting biases, but which do not include a structured tool that can be applied by users (e.g. the 2011 Cochrane Handbook chapter on reporting biases[22]);
- tools developed or modified for use in one particular systematic review;

- tools designed to appraise published systematic reviews, such as the ROBIS tool[23] or AMSTAR[24];
- articles that focus on the development or evaluation of statistical methods to detect or adjust for reporting biases, as these have been reviewed elsewhere[18-20].

Search methods

On 9 February 2017, one author (MJP) searched for potentially relevant records in Ovid MEDLINE (January 1946 to February 2017), Ovid EMBASE (January 1980 to February 2017), and Ovid PsycINFO (January 1806 to February 2017). The search strategies included terms relating to reporting bias, which were combined with a search string used previously by Whiting et al. to identify risk of bias/quality assessment tools[17] (see full Boolean search strategies in online supplementary table S1).

To capture any tools not published by formal academic publishers, we searched Google Scholar using the phrase "reporting bias tool OR risk of bias". One author (MJP) screened the titles of the first 300 records, as recommended by Haddaway et al.[25]. To capture any papers that may have been missed by all searches, one author (MJP) screened the references of included articles. In April 2017, the same author emailed the list of included tools to 15 individuals with expertise in reporting biases and risk of bias assessment, and asked if they were aware of any other tools we had not identified.

Study selection and data collection

One author (MJP) screened all titles and abstracts retrieved by the searches. The same author screened any full text articles retrieved. One author (MJP) collected data from included papers using a standardised data collection form. The following data on included tools were collected:

- type of tool (scale, checklist, or domain-based tool);
- types of reporting bias addressed by the tool;

- level of assessment (i.e. whether users direct assessments at the synthesis or at the individual studies included in the synthesis);
- whether the tool is designed for general use (generic) or targets specific study designs or topic areas (specific);
- items included in the tool;
- how items within the tool are rated;
- methods used to develop the tool (e.g. Delphi study, expert consensus meeting);
- availability of guidance to assist with completion of the tool (e.g. guidance manual).

The following data from studies evaluating measurement properties of an included tool were collected:

- tool evaluated;
- measurement properties evaluated (e.g. inter-rater agreement);
- number of syntheses/studies evaluated;
- publication year of syntheses/studies evaluated;
- areas of health care addressed by syntheses/studies evaluated;
- number of assessors;
- estimate (and precision) of psychometric statistics (e.g. weighted kappa).

Data analysis

We summarised the characteristics of included tools in tables. We calculated the median (interquartile range (IQR)) number of items across all tools, and tabulated the frequency of different criteria used in tools to denote a judgement of “high” risk of reporting bias. We summarised estimates of psychometric statistics, such as weighted kappa to estimate inter-rater agreement[26], by reporting the range of values across studies. For studies reporting weighted kappa, we categorised agreement according to the system proposed by Landis et al.[27], as poor (0.00), slight (0.01-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80), or almost perfect (0.81-1.00).

RESULTS

In total, 5,554 records were identified from the searches, of which we retrieved 165 for full text screening (Figure 1). The inclusion criteria were met by 42 reports summarising 18 tools (Table 1) and 17 studies evaluating the measurement properties of tools[3 4 21 28-66]. A list of excluded papers is presented in online supplementary Table S2. No additional tools were identified by the 15 experts contacted.

Table 1. List of included tools

Article ID	Tool	Scope of tool	Types of reporting biases assessed			Level of assessment[a]
			Selective publication	**Selective non-reporting**	**Selection of the reported result**	
Balshem 2013[28]	AHRQ outcome and analysis reporting bias framework	Reporting bias only		✓	✓	Specific outcome/ result in a study
Berkman 2013[29]	AHRQ tool for evaluating the risk of reporting bias	Reporting bias only	✓	✓		Specific synthesis of studies
Downes 2016[30]	AXIS tool (Appraisal tool for Cross-Sectional Studies)	Multiple sources of bias		✓		Study
Downs 1998[31]	Downs-Black tool	Multiple sources of bias			✓	Study
Guyatt 2011[33-37]	GRADE	Multiple sources of bias	✓	✓		Specific synthesis of studies
Hayden 2013[38]	QUIPS (Quality In Prognosis Studies) tool	Multiple sources of bias		✓		Study
Higgins 2008[21 39 40]	Cochrane risk of bias tool for randomized trials (RoB 1.0)	Multiple sources of bias		✓	✓	Study
Higgins 2016[41 42]	RoB 2.0 revised tool for assessing risk of bias in randomized trials	Multiple sources of bias			✓	Specific result in a study
Hoojimans 2014[43]	SYRCLE's RoB tool (SYstematic Review Centre for Laboratory animal Experimentation)	Multiple sources of bias		✓	✓	Study
Kim 2013[44]	RoBANS (Risk of Bias Assessment Tool for Nonrandomized Studies)	Multiple sources of bias		✓	✓	Study
Kirkham	ORBIT-I (Outcome Reporting Bias In Trials)	Reporting bias		✓		Specific outcome

Article ID	Tool	Scope of tool	Types of reporting biases assessed			Level of assessment[a]
			Selective publication	Selective non-reporting	Selection of the reported result	
2010[3 32]	classification system for benefit outcomes	only				in a study
Meader 2014[45 46]	SAQAT (Semi-Automated Quality Assessment Tool)	Multiple sources of bias	✓	✓		Specific synthesis of studies
Reid 2015[47]	Selective reporting bias algorithm	Reporting bias only		✓	✓	Study
Saini 2014[48]	ORBIT-II (Outcome Reporting Bias In Trials) classification system for harm outcomes	Reporting bias only		✓		Specific outcome/ result in a study
Salanti 2014[49 50]	Framework for evaluating the quality of evidence from a network meta-analysis	Multiple sources of bias	✓	✓		Specific synthesis of studies
Sterne 2016[4]	ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool	Multiple sources of bias			✓	Specific result in a study
Viswanathan 2012[51]	RTI Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures	Multiple sources of bias		✓		Study
Viswanathan 2013[52]	RTI Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures	Multiple sources of bias		✓		Study

[a] Level of assessment classified as: "study" when assessments are directed at a study as a whole (e.g. tool used to assess whether *any* outcomes in a study were not reported); "specific outcome/result in a study" when assessments are directed at a specific outcome or result within a study (e.g. tools used to assess whether a particular outcome, such as pain, was not reported) or; "specific synthesis of studies" when assessments are directed at a specific synthesis (e.g. tool used to assess whether a particular synthesis, such as a meta-analysis of pain, is missing unpublished studies).

General characteristics of included tools

Nearly all of the included tools (16/18 [89%]) were domain-based, where users judge risk of bias or quality within specific domains (Table 2; individual characteristics of each tool are presented in online supplementary Table S3). All tools were designed for generic rather than specific use. Five tools focused solely on the risk of reporting biases[3 28 29 47 48]; the remainder addressed reporting biases and other sources of bias/methodological quality (e.g. problems with randomization, lack of blinding). Half of the tools (9/18 [50%]) addressed only one type of reporting bias (e.g. bias due to selective non-reporting only). Tools varied in regard to the study design that they assessed (i.e. randomized trial, non-randomized study of an intervention, laboratory animal experiment). The publication year of the tools ranged from 1998 to 2016 (the earliest was the Downs-Black tool[31], a 27-item tool assessing multiple sources of bias, one of which focuses on risk of bias in selection of the reported result).

Assessments for half of the tools (9/18 [50%]) are directed at an individual study (e.g. tool is used to assess whether *any outcomes in a study* were not reported). In 5/18 (28%) tools, assessments are directed at a specific outcome or result within a study (e.g. tool is used to assess whether *a particular outcome in a study*, such as pain, was not reported). In a few tools (4/18 [22%]), assessments are directed at a specific synthesis (e.g. tool is used to assess whether *a particular synthesis*, such as a meta-analysis of studies examining pain as an outcome, is missing unpublished studies).

The content of the included tools was informed by various sources of data. The most common included a literature review of items used in existing tools or a literature review of empirical evidence of bias (9/18 [50%]), ideas generated at an expert consensus meeting (8/18 [44%]) and pilot feedback on a preliminary version of the tool (7/18 [39%]). The most common type of guidance available for the tools was a brief annotation per item/response option (9/18 [50%]). A detailed guidance manual is available for four (22%) tools.

Table 2. Summary of general characteristics of included tools

Characteristic	Summary data (n = 18 tools)
Type of tool	
Domain-based	16 (89%)
Checklist	1 (6%)
Scale	1 (6%)
Scope of tool	
Assessment of reporting bias only	5 (28%)
Assessment of multiple sources of bias/quality	13 (72%)
Types of reporting bias assessed	
Bias due to selective publication only	0 (0%)
Bias due to selective non-reporting only	6 (33%)
Bias in selection of the reported result only	3 (17%)
Bias due to selective publication and bias due to selective non-reporting	4 (22%)
Bias due to selective non-reporting and bias in selection of the reported result	5 (28%)
Total number of items in the tool	7 (5-13)
Number of items relevant to risk of reporting bias	1 (1-2)
Number of response options for risk of reporting bias judgement	3 (3-3)
Types of study designs to which the tool applies	
Randomized trials only	5 (28%)
Systematic reviews only	3 (17%)
Non-randomized studies of interventions only	2 (11%)
Randomized trials and non-randomized studies of interventions	2 (11%)
Non-randomized studies of interventions or exposures	2 (11%)
Other (cross-sectional studies, animal studies, network meta-analyses, prognosis studies)	4 (22%)
Level of assessment of risk of reporting bias	
Study as a whole	9 (50%)
Specific outcome/result in a study	5 (28%)
Specific synthesis of studies	4 (22%)
Data sources used to inform tool content[a]	
Literature review (e.g. of items in existing tools, or empirical evidence)	9 (50%)
Ideas generated at expert consensus meeting	8 (44%)
Pilot feedback on preliminary version of the tool	7 (39%)

Characteristic	Summary data (n = 18 tools)
Data from psychometric or cognitive testing[b]	5 (28%)
Other (e.g. adaptation of existing tool)	5 (28%)
Delphi study responses	2 (11%)
No methods stated	2 (11%)
Guidance available	
Brief annotation per item/response option	9 (50%)
Detailed guidance manual	4 (22%)
Worked example for each response option	2 (11%)
Detailed annotation per item/response option	1 (6%)
None	2 (11%)

Summary data given as number (percent) or median (IQR).
[a]The percentages in this category do not sum to 100% since the development of some tools was informed by multiple data sources.
[b]Psychometric testing includes any evaluation of the measurement properties (e.g. construct validity, inter-rater reliability, test-retest reliability) of a draft version of the tool. Cognitive testing includes use of qualitative methods (e.g. interview) to explore whether assessors who are using the tool for the first time were interpreting the tool and guidance as intended.

Tool content

Four tools include items for assessing risk of bias due to both selective publication and selective non-reporting[29 33 45 49]. One of these tools (the AHRQ tool for evaluating the risk of reporting bias[29]) directs users to assess a particular synthesis, where a single risk of bias judgement is made based on information about unpublished studies and underreported outcomes. In the other three tools (the GRADE framework, and two others which are based on GRADE[33 45 49]), the different sources of reporting bias are assessed in separate domains (bias due to selective non-reporting is considered in a “study limitations (risk of bias)” domain, while bias due to selective publication is considered in a “publication bias” domain).

Five tools[21 28 43 44 47] guide users to assess risk of bias due to both selective non-reporting and selection of the reported result (that is, problems with outcomes/results that *are not* reported and those that *are* reported, respectively). Four of these tools, which include the Cochrane risk of bias tool for randomized trials[21] and three others which are based on the Cochrane tool[43 44 47], direct assessments at the study level. That is, a whole study is rated at “high” risk of reporting bias if *any* outcome/result in the study has been omitted, or fully reported, on the basis of the findings.

Some of the tools designed to assess the risk of bias due to selective non-reporting ask users to assess, for particular outcomes of interest, whether the outcome was not reported or only partially reported in the study on the basis of its results (e.g. ORBIT tools[3 48], the AHRQ outcome reporting bias framework[28], and GRADE[34]). This allows users to perform multiple outcome-level assessments of the risk of reporting bias (rather than one assessment for the study as a whole). In total, 15 tools include a mechanism for assessing risk of bias due to selective non-reporting in studies, but assessing the corresponding risk of bias in a synthesis that is missing the non-reported outcomes is not within the scope of 11 of these tools [3 21 28 30 38 43 44 47 48 51 52].

A variety of criteria are used in existing tools to inform a judgement of “high” risk of bias due to selective publication (Table 3), selective non-reporting (Table 4), and selection of the reported result (Table 5) (more detail is provided in online supplementary Table S4). In the four tools with an assessment of risk of bias due to selective publication, “high” risk criteria include evidence of funnel plot asymmetry, discrepancies between published and unpublished studies, use of non-comprehensive searches, and presence of small, “positive” studies with for-profit interest (Table 3). However, not all of these criteria appear in all tools (only evidence of funnel plot asymmetry does), and the relative weight assigned to each criterion in the overall risk of reporting bias judgement is clear for only one tool (the Semi-Automated Quality Assessment Tool (SAQAT)[45 46]).

All 15 tools with an assessment of the risk of bias due to selective non-reporting suggest that the risk of bias is “high” when it is clear that an outcome was measured but no results were reported (Table

4). Fewer of these tools (n=8 [53%]) also recommend a "high" risk judgement when results for an outcome are partially reported (e.g. it is stated that the result was non-significant, but no effect estimate or summary statistics are presented).

The eight tools that include an assessment of the risk of bias in selection of the reported result recommend various criteria for a "high" risk judgement (Table 5). These include when some outcomes that were not pre-specified are added post-hoc (in 4 [50%] tools), or when it is likely that the reported result for a particular outcome has been selected, on the basis of the findings, from amongst multiple outcome measurements or analyses within the outcome domain (in 2 [25%] tools).

Table 3. Criteria used in existing tools to inform a judgement of "high" risk of bias due to selective publication

"High" risk of bias criteria proposed in existing tools	AHRQ RRB	GRADE	SAQAT	NMA-Quality	Total n (%)
Assessment directed at a specific synthesis (e.g. meta-analysis)					
Evidence of funnel plot asymmetry (based on visual inspection of funnel plot or statistical test for funnel plot asymmetry)	✓	✓	✓	✓	4 (100)
Smaller studies tend to demonstrate more favourable results (based on visual assessment, without funnel plot)	✓				1 (25)
Clinical decision would differ for estimates from a fixed-effect versus a random-effects model, because the findings from a fixed-effect model are closer to the null	✓				1 (25)
Substantial heterogeneity in the meta-analysis cannot be explained by some clinical or methodological factor	✓				1 (25)
At least one study is affected by non-publication or non-accessibility	✓				1 (25)
Presence of small (often "positive") studies with for-profit interest in the synthesis		✓		✓	2 (50)
Presence of early studies (i.e. set of small, "positive" trials addressing a novel therapy) in the synthesis		✓		✓	2 (50)
Discrepancy in findings between published and unpublished trials		✓	✓	✓	3 (75)
Search strategies were not comprehensive		✓	✓	✓	3 (75)
Methods to identify all available evidence were not comprehensive		✓		✓	2 (50)
Grey literature were not searched			✓		1 (25)
Restrictions to study selection on the basis of language were applied			✓		1 (25)
Industry influence may apply to studies included in the synthesis			✓		1 (25)

AHRQ RRB = AHRQ tool for evaluating the risk of reporting bias[29]; GRADE = GRADE rating of quality of evidence[34-37]; NMA-Quality = Framework for evaluating the quality of evidence from a network meta-analysis[49]; SAQAT = Semi-Automated Quality Assessment Tool[45 46].

Table 4. Criteria used in existing tools to inform a judgement of "high" risk of bias due to selective non-reporting

"High" risk of bias criteria proposed in existing tools	AHRQ ORB	AHRQ RRB	AXIS	GRADE	QUIPS	RoB 1.0	SYRCLE RoB	RoBANS	ORBIT-I	SAQAT	Reid	ORBIT-II	NMA-Quality	RTI 2012	RTI 2013	Total n (%)
Assessment directed at study as a whole																
One or more outcomes of interest were clearly measured, but no results were reported			✓		✓	✓	✓	✓			✓			✓	✓	8 (53)
One or more outcomes of interest are reported incompletely so that they cannot be entered in a meta-analysis						✓		✓								2 (13)
The study report fails to include results for a key outcome that would be expected to have been reported for such a study						✓	✓	✓						✓	✓	5 (33)
Assessment directed at a specific outcome																
Particular outcome clearly measured, but no results were reported	✓	✓		✓					✓			✓	✓			6 (40)
Particular outcome of interest is reported incompletely so that it cannot be entered in a meta-analysis (typically stating only that P>0.05).	✓	✓		✓					✓			✓	✓			6 (40)

"High" risk of bias criteria proposed in existing tools	AHRQ ORB	AHRQ RRB	AXIS	GRADE	QUIPS	RoB 1.0	SYRCLE RoB	RoBANS	ORBIT-I	SAQAT	Reid	ORBIT-II	NMA-Quality	RTI 2012	RTI 2013	Total n (%)
Judgment says particular outcome is likely to have been measured and analysed but not reported on the basis of its results	✓	✓		✓					✓			✓	✓			6 (40)
Composite outcomes are presented without the individual component outcomes				✓								✓				2 (13)
Result reported globally across all groups												✓				1 (7)
Result reported for some groups only												✓				1 (7)
Data were not reported consistently for the outcome of interest										✓						1 (7)
Assessment directed at a specific synthesis																
Selective non-reporting suspected in a number of included studies		✓		✓						✓			✓			4 (27)

AHRQ ORB = AHRQ outcome and analysis reporting bias framework[28]; AHRQ RRB = AHRQ tool for evaluating the risk of reporting bias[29]; AXIS = Appraisal tool for Cross-Sectional Studies[30]; GRADE = GRADE rating of quality of evidence[34-37]; NMA-Quality = Framework for evaluating the quality of evidence from a network meta-analysis[49]; ORBIT-I = Outcome Reporting Bias In Trials classification system for benefit outcomes[3 32]; ORBIT-II = Outcome Reporting Bias In Trials classification system for harm outcomes[48]; QUIPS = Quality In Prognosis Studies tool[38]; Reid = Reid et al. selective reporting bias algorithm[47]; RoB 1.0 = Cochrane risk of bias tool for randomized trials[21 39 40]; RoBANS = Risk of Bias Assessment Tool for Nonrandomized Studies[44]; RTI 2012 = RTI Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures[51]; RTI 2013 = RTI Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures[52]; SAQAT = Semi-Automated Quality Assessment Tool[45 46]; SYRCLE RoB = SYstematic Review Centre for Laboratory animal Experimentation risk of bias tool[43].

Table 5. Criteria used in existing tools to inform a judgement of "high" risk of bias in selection of the reported result

"High" risk of bias criteria proposed in existing tools	AHRQ ORB	Downs-Black	RoB 1.0	RoB 2.0	SYRCLE RoB	RoBANS	Reid	ROBINS-I	Total n (%)
Assessment directed at study as a whole									
One or more reported outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse event)			✓		✓	✓	✓		4 (50)
One or more outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified			✓		✓				2 (15)
One or more retrospective, unplanned, subgroup analysis was reported		✓							1 (13)
Any analyses that had not been planned at the outset of the study were not clearly indicated		✓							1 (13)
Assessment directed at a specific outcome/result									
Particular outcome was not pre-specified but results were reported	✓								1 (13)
Reported result for a particular outcome is likely to have been selected, on the basis of the findings, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain				✓				✓	2 (25)
Reported result for a particular outcome is likely to have been selected, on the basis of the findings, from multiple analyses of the data				✓				✓	2 (25)
Reported result for a particular outcome is likely to have been selected, on the basis of the findings, from different subgroups								✓	1 (13)

AHRQ ORB = AHRQ outcome and analysis reporting bias framework[28]; Downs-Black = Downs-Black tool[31]; Reid = Reid et al. selective reporting bias algorithm[47]; RoB 1.0 = Cochrane risk of bias tool for randomized trials[21 39 40]; RoB 2.0 = Revised tool for assessing risk of bias in randomized trials[41 42]; RoBANS = Risk of Bias Assessment Tool for Nonrandomized Studies[44]; ROBINS-I = Risk Of Bias In Non-randomized Studies of Interventions tool[4]; SYRCLE RoB = SYstematic Review Centre for Laboratory animal Experimentation risk of bias tool[43].

General characteristics of studies evaluating measurement properties of included tools

Despite identifying 17 studies that evaluated measurement properties of an included tool, psychometric statistics for the risk of reporting bias component were available only from 12 studies[43 44 54-60 62 64 66] (the other five studies include only data on properties of the multi-dimensional tool as a whole[31 53 61 63 65]) (online supplementary Table S5). Nearly all 12 studies (11 [92%]) evaluated inter-rater agreement between two assessors; eight of these studies reported weighted kappa (κ) values, but only two described the weighting scheme[55 62]. Eleven studies[43 44 54-60 64 66] evaluated the measurement properties of tools for assessing risk of bias in a study due to selective non-reporting or risk of bias in selection of the reported result; in these 11 studies, a median of 40 (IQR 32-109) studies were assessed. One study[62] evaluated a tool for assessing risk of bias in a synthesis due to selective publication, in which 44 syntheses were assessed. In the studies evaluating inter-rater agreement, all involved two assessors.

Results of evaluation studies

Five studies[54 56-58 60] included data on the inter-rater agreement of assessments of risk of bias due to selective non-reporting using the Cochrane risk of bias tool for randomized trials[21] (Table 6). Weighted kappa (κ) values in four studies[54 56-58] ranged from 0.13 to 0.50 (sample size ranged from 87 to 163 studies), suggesting slight to moderate agreement[27]. In the other study[60], the percent agreement in selective non-reporting assessments in trials that were included in two different Cochrane reviews was low (43% of judgements were in agreement). Two other studies found that inter-rater agreement of selective non-reporting assessments were substantial for SYRCLE's RoB tool (κ = 0.62, n = 32)[43], but poor for the RoBANS tool (κ = 0, n = 39)[44]. There was substantial agreement between raters in the assessment of risk of bias due to selective publication using the SAQAT (κ = 0.63, n = 29)[62]. The inter-rater agreement of assessments of risk of bias in selection of the reported result using the ROBINS-I tool[4] was moderate for NRSI included in a review of the effect of cyclooxygenase-2 (COX-2) inhibitors on cardiovascular events (κ = 0.45, n = 21), and substantial for

NRSI included in a review of the effect of thiazolidinediones on cardiovascular events (κ = 0.78, n = 16)[55].

Table 6. Reported measurement properties of tools with an assessment of the risk of reporting bias

Study ID	Tool	Measurement property	Sample size	Areas of health care addressed	Weighted kappa (95% CI)	Weighting scheme	Interpretation of kappa[a]
Armijo-Olivo 2014[54]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between two external reviewers)	87	Musculoskeletal, cardiorespiratory, neurological, and gynaecological conditions	0.5 (CI not reported)	Not described	Moderate agreement
Armijo-Olivo 2014[54]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between two external reviewers and Cochrane reviewers)	87	See above	0.13 (CI not reported)	Not described	Slight agreement
Hartling 2009[56]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting	163	Child health	0.13 (95% CI -0.05 to 0.31)	Not described	Slight agreement
Hartling 2011[57]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting	107	Asthma	0.4 (95% CI 0.14 to 0.67)	Not described	Fair agreement
Hartling 2012[58] [59]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between two reviewers, all trials)	124	Varied	0.27 (95% CI 0.06 to 0.49)	Not described	Fair agreement
Hartling 2012[58] [59]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between pairs of reviewers across different centres, all trials)	30	Varied	0.08 (95% CI -0.09 to 0.26)	Not described	Slight agreement
Jordan 2017[60]	RoB 1.0	Inter-rater agreement of assessments of risk of bias due to selective non-reporting (between judgements of trials appearing in two SRs)	28	Subfertility	Not reported[b]	Not applicable	Not applicable
Vale 2013[66]	RoB 1.0	Agreement between selective non-reporting assessments performed using published article only versus published article and data collected during the individual participant data process	95	Cancer pain	Not reported[b]	Not applicable	Not applicable
Hoojimans	SYRCLE RoB	Inter-rater agreement of assessments of risk of	32	Animal studies (not	0.62 (CI not	Not	Substantial

Study ID	Tool	Measurement property	Sample size	Areas of health care addressed	Weighted kappa (95% CI)	Weighting scheme	Interpretation of kappa[a]
2014[43]		bias due to selective non-reporting		specified)	reported)	described	agreement
Kim 2013[44]	RoBANS	Inter-rater agreement of assessments of risk of bias due to selective non-reporting	39	Depression, myocardial infarction, post-partum hemorrhage, chronic non-cancer pain	0 (CI not reported)	Not described	Poor agreement
Llewellyn 2015[62]	SAQAT	Inter-rater agreement of assessments of risk of bias due to selective publication (between two SAQAT raters)	29	Varied	0.63 (95% CI 0.17 to 1)	Quadratic	Substantial agreement
Llewellyn 2015[62]	SAQAT	Inter-rater agreement of assessments of risk of bias due to selective publication (between one rater using SAQAT and one using the standard GRADE approach)	15	Varied	Not reported[b]	Not applicable	Not applicable
Norris 2012[64]	ORBIT-I	Inter-rater agreement of ORBIT-I classifications of risk of bias due to selective non-reporting	40	Varied	Not calculated, as too little variation in judgements	Not applicable	Not applicable
Bilandzic 2016[55]	ROBINS-I	Inter-rater agreement of assessments of risk of bias in selection of the reported result	16	Thiazolidinediones and cardiovascular events	0.78 (CI not reported)	Linear	Substantial agreement
Bilandzic 2016[55]	ROBINS-I	Inter-rater agreement of assessments of risk of bias in selection of the reported result	21	COX-2 inhibitors and cardiovascular events	0.45 (CI not reported)	Linear	Moderate agreement

[a] Interpretation of kappa based on categorisation system defined by Landis et al.[27]. [b] Data presented as percent agreement, not weighted kappa. ORBIT-I = Outcome Reporting Bias In Trials classification system for benefit outcomes[3 32]; RoB 1.0 = Cochrane risk of bias tool for randomized trials[21 39 40]; RoBANS = Risk of Bias Assessment Tool for Nonrandomized Studies[44]; ROBINS-I = Risk Of Bias In Non-randomized Studies of Interventions tool[4]; SAQAT = Semi-Automated Quality Assessment Tool[45 46]; SRs = systematic reviews; SYRCLE RoB = SYstematic Review Centre for Laboratory animal Experimentation risk of bias tool[43].

DISCUSSION

From a systematic search of the literature, we identified 18 tools designed for use by individuals performing evidence syntheses to assess risk of reporting biases in the included studies or in their synthesis of studies. The tools varied with regard to the type of reporting bias assessed (e.g. bias due to selective publication, bias due to selective non-reporting), and the level of assessment (e.g. for the study as a whole, a particular outcome within a study, or a particular synthesis of studies). Various criteria are used across tools to designate a synthesis as being at "high" risk of bias due to selective publication (e.g. evidence of funnel plot asymmetry, use of non-comprehensive searches). However, the relative weight assigned to each criterion in the overall judgement is not clear for most of these tools. Tools for assessing risk of bias due to selective non-reporting guide users to assess a study, or an outcome within a study, as "high" risk of bias if no results are reported for an outcome. However, assessing the corresponding risk of bias in a synthesis that is missing the non-reported outcomes is outside the scope of most of these tools. Inter-rater agreement estimates were available for five tools[4 21 43 44 62], and ranged from poor to substantial; however the sample sizes of most evaluations were small, and few described the weighting scheme used to calculate kappa.

Strengths and limitations

There are several strengths of this research. Methods were conducted in accordance with a systematic review protocol (https://osf.io/9ea22/). Published articles were identified by searching several relevant databases using a search string developed in conjunction with an information specialist[17], and by contacting experts to identify tools missed by the search. Detailed information on the content and measurement properties of existing tools was collected, providing readers with pertinent information to help decide which tools to use in future reviews. However, the findings need to be considered in light of some limitations. Screening of articles and data collection were performed by one author only. It is therefore possible that some relevant articles were missed, or that errors in data collection were made. The search for unpublished tools was not comprehensive (only Google Scholar was searched), so it is possible that other tools for assessing risk of reporting biases exist. Further, restricting the search to articles in English was done to expedite the review process, but may have resulted in loss of information about tools written in other languages, and additional evidence on measurement properties of tools.

Comparison with other studies

Other systematic reviews of risk of bias tools[12-17] have restricted inclusion to tools developed for particular study designs (e.g. randomized trials, diagnostic test accuracy studies), where the authors recorded all the sources of bias addressed. A different approach was taken in the current review, where all tools (regardless of study design) that address a particular source of bias were examined. By focusing on one source of bias only, the analysis of included items and criteria for risk of bias judgements was more detailed than that recorded previously. Some of the existing reviews of tools[15] considered tools that were developed or modified in the context of a specific systematic review. However, such tools were excluded from the current review as they are unlikely to have been developed systematically[15 67], and are difficult to find (all systematic reviews conducted during a particular period would need to have been examined for the search to be considered exhaustive).

Explanations and implications

Of the 18 tools identified, only four (22%) included a mechanism for assessing risk of bias due to selective publication, which is the type of reporting bias that has been investigated by methodologists most often[2]. This is perhaps unsurprising given that hundreds of statistical methods to “detect” or “adjust” for bias due to selective publication have been developed[18]. These statistical methods may be considered by methodologists and systematic reviewers as the tools of choice for assessing this type of bias. However, application of these statistical methods without considering other factors (e.g. existence of registered but unpublished studies, conflicts of interest that may influence investigators to not disseminate studies with unfavourable results) is not sufficiently comprehensive, and could lead to incorrect conclusions about the risk of bias due to selective publication. Further, there are many limitations of these statistical approaches, in terms of their underlying assumptions, statistical power, which is often low because most meta-analyses include few studies[7], and the need for specialist statistical software to apply them[19 68]. These factors may have limited their use in practice, and potentially explain why a large number of systematic reviewers currently ignore the risk of bias due to selective publication[7-9 69].

Our analysis suggests that the factors that need to be considered to assess risk of reporting biases adequately (e.g. comprehensiveness of the search, amount of data missing from the synthesis due to unpublished studies and underreported outcomes) are fragmented. A similar problem was occurring a decade ago with the assessment of risk of bias in randomized trials. Some authors assessed only problems with randomization, while others focused on whether trials were not “double blinded”, or had any missing participant data[70]. It was not until all the important bias domains were brought together into a structured, domain-based tool to assess the risk of bias in randomized trials[21], that systematic reviewers started to consider risk of bias in trials comprehensively. A similar initiative to link all the components needed to judge the risk of reporting biases into a comprehensive new tool may improve the credibility of evidence syntheses.

In particular, there is an emergent need for a new tool to assess the risk that a synthesis is affected by reporting biases. This tool could guide users to consider risk of bias in a synthesis due to both selective publication and selective non-reporting, given that both practices lead to the same consequence: evidence missing from the synthesis[11]. Such a tool would complement recently developed tools for assessing risk of bias within studies (RoB 2.0[41] and ROBINS-I[4]) which include a domain for assessing the risk of bias in selection of the reported result, but no mechanism to assess risk of bias due to selective non-reporting. Careful thought would need to be given as to how to weigh up various pieces of information underpinning the risk of bias judgement. For example, users will need guidance on how evidence of known, unpublished studies (as identified from trial registries, protocols or regulatory documents) should be considered alongside evidence that is more speculative (e.g. funnel plots suggesting that studies may be missing). Further, guidance for the tool will need to emphasise the value of seeking documents other than published journal articles (e.g. protocols) to inform risk of bias judgements. Preparation of a detailed guidance manual may enhance the usability of the tool, minimise misinterpretation and increase reliability in assessments. Once developed, evaluations of the measurement properties of the tool, such as inter-rater agreement and construct validity, should be conducted to explore whether modifications to the tool are necessary.

Conclusions

There are several limitations of existing tools for assessing risk of reporting biases in studies or syntheses of studies, in terms of their scope, guidance for reaching risk of bias judgements, and measurement properties. Development and evaluation of a new, comprehensive tool, could help overcome present limitations.

Acknowledgments

Not applicable.

Competing Interests

We have read the journal's policy and have the following competing interests: JPTH led or participated in the development of four of the included tools (the current Cochrane risk of bias tool for randomized trials, the RoB 2.0 tool for assessing risk of bias in randomized trials, the ROBINS-I tool for assessing risk of bias in non-randomized studies of interventions, and the framework for assessing quality of evidence from a network meta-analysis). MJP participated in the development of one of the included tools (the RoB 2.0 tool for assessing risk of bias in randomized trials). All authors are participating in the development of a new tool for assessing risk of reporting biases in systematic reviews.

Author Contributions

MJP conceived and designed the study, collected data, analysed the data, and wrote the first draft of the article. JM and JPTH provided input on the study design and contributed to revisions of the article. All authors approved the final version of the submitted article.

Data sharing statement

The study protocol, data collection form, and the raw data and statistical analysis code for this study are available on the Open Science Framework: https://osf.io/3jdaa/

Funding

There was no direct funding for this study. MJP is supported by an Australian National Health and Medical Research Council (NHMRC) Early Career Fellowship (1088535). JEM is supported by a NHMRC Australian Public Health Fellowship (1072366). JPTH is funded in part by Cancer Research UK Programme Grant C18281/A19169; is a member of the MRC Integrative Epidemiology Unit at the University of Bristol, which is supported by the UK Medical Research Council and the University of Bristol (grant MC_UU_12013/9); and is a member of the MRC ConDuCT-II Hub (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures; grant MR/K025643/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Chan A-W, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *The Lancet* 2014;383(9913):257-66.

2. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. *Health Technol Assess* 2010;14:8.

3. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.

4. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.

5. Schmucker C, Schell LK, Portalupi S, et al. Extent of non-publication in cohorts of studies approved by research ethics committees or included in trial registries. *PLoS One* 2014;9(12):e114023.

6. Jones CW, Keil LG, Holland WC, et al. Comparison of registered and published outcomes in randomized controlled trials: a systematic review. *BMC Med* 2015;13:282.

7. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med* 2016;13(5):e1002028.

8. Koletsi D, Valla K, Fleming PS, et al. Assessment of publication bias required improvement in oral health systematic reviews. *J Clin Epidemiol* 2016;76:118-24

9. Hedin RJ, Umberham BA, Detweiler BN, et al. Publication Bias and Nonreporting Found in Majority of Systematic Reviews and Meta-analyses in Anesthesiology Journals. *Anesth Analg* 2016;123(4):1018-25.

10. Ziai H, Zhang R, Chan AW, et al. Search for unpublished data by systematic reviewers: an audit. *BMJ open* 2017;7(10):e017737.

11. Page MJ, Higgins JPT. Rethinking the assessment of risk of bias due to selective reporting: a cross-sectional study. *Systematic reviews* 2016;5(1):108.

12. Moher D, Jadad AR, Nichol G, et al. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;16(1):62-73.

13. Armijo Olivo S, Macedo LG, Gadotti IC, et al. Scales to assess the quality of randomized controlled trials: a systematic review. *Phys Ther* 2008;88(2):156-75.

14. Bai A, Shukla VK, Bak G, et al. Quality Assessment Tools Project Report. Ottawa: Canadian Agency for Drugs and Technologies in Health, 2012.

15. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007;36(3):666-76.

16. Whiting P, Rutjes AW, Dinnes J, et al. A systematic review finds that diagnostic reviews fail to incorporate quality despite available tools. *J Clin Epidemiol* 2005;58(1):1-12.

17. Whiting P, Davies P, Savovic J, et al. Evidence to inform the development of ROBIS, a new tool to assess the risk of bias in systematic reviews, September 2013. Available from https://www.researchgate.net/publication/303312018_Evidence_to_inform_the_development_of_ROBIS_a_new_tool_to_assess_the_risk_of_bias_in_systematic_reviews [accessed 1 August 2017].

18. Mueller KF, Meerpohl JJ, Briel M, et al. Methods for detecting, quantifying and adjusting for dissemination bias in meta-analysis are described. *J Clin Epidemiol* 2016;80:25-33.

19. Jin ZC, Zhou XH, He J. Statistical methods for dealing with publication bias in meta-analysis. *Stat Med* 2015;34(2):343-60.

20. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002.

21. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

22. Sterne JAC, Egger M, Moher D. Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions Version 510 [updated March 2011] 2011.

23. Whiting P, Savovic J, Higgins JP, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;69:225-34.

24. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.

25. Haddaway NR, Collins AM, Coughlin D, et al. The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. *PLoS One* 2015;10(9):e0138237.

26. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20(1):37-46.

27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-74.

28. Balshem H, Stevens A, Ansari M, et al. Finding grey literature evidence and assessing for outcome and analysis reporting biases when comparing medical interventions: AHRQ and the Effective Health Care Program. (Prepared by the Oregon Health and Science University and the University of Ottawa Evidence-based Practice Centers under Contract Nos. 290-2007-10057-I and 290-2007-10059-I.) AHRQ Publication No. 13(14)-EHC096-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

29. Berkman ND, Lohr KN, Ansari M, et al. Chapter 15 Appendix A: A Tool for Evaluating the Risk of Reporting Bias (in Chapter 15: Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update). Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I). AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm

30. Downes MJ, Brennan ML, Williams HC, et al. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ open* 2016;6:e011458.

31. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.

32. Dwan K, Gamble C, Kolamunnage-Dona R, et al. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials* 2010;11:52.

33. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.

34. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol* 2011;64(4):407-15.

35. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64(12):1277-82.

36. Schünemann H, Brożek J, Guyatt G, et al. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. [Updated October 2013]. Available from http://gdt.guidelinedevelopment.org/app/handbook/handbook.html.

37. Santesso N, Carrasco-Labra A, Langendam M, et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. *J Clin Epidemiol* 2016

38. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.

39. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons 2008:187-241.

40. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org/.

41. Higgins JPT, Savović J, Page MJ, et al. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0), Version 20 October 2016. Available from https://sites.google.com/site/riskofbiastool/ [accessed 19 September 2017].

42. Higgins JPT, Sterne JAC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials. *Cochrane Methods Cochrane Database of Systematic Reviews* 2016;10(Suppl 1):29-31.

43. Hooijmans CR, Rovers MM, de Vries RBM, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.

44. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013;66(4):408-14.

45. Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic reviews* 2014;3(1):82.

46. Stewart GB, Higgins JP, Schunemann H, et al. The use of Bayesian networks to assess the quality of evidence from research synthesis: 1. *PLoS One* 2015;10(3):e0114497.

47. Reid EK, Tejani AM, Huan LN, et al. Managing the incidence of selective reporting bias: a survey of Cochrane review groups. *Systematic reviews* 2015;4:85.

48. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ* 2014;349:g6501.

49. Salanti G, Giovane CD, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9(7):e99682.

50. Higgins JP, Del Giovane C, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *Value Health* 2014;17(7):A324.

51. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* 2012;65(2):163-78.

52. Viswanathan M, Berkman ND, Dryden DM, et al. AHRQ Methods for Effective Health Care. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or

Exposures: Further Development of the RTI Item Bank. Rockville (MD): Agency for Healthcare Research and Quality (US) 2013.

53. Armijo-Olivo S, Stiles CR, Hagen NA, et al. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract* 2012;18(1):12-8.

54. Armijo-Olivo S, Ospina M, da Costa BR, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. *PLoS One* 2014;9(5):e96920.

55. Bilandzic A, Fitzpatrick T, Rosella L, et al. Risk of Bias in Systematic Reviews of Non-Randomized Studies of Adverse Cardiovascular Effects of Thiazolidinediones and Cyclooxygenase-2 Inhibitors: Application of a New Cochrane Risk of Bias Tool. *PLoS Med* 2016;13(4):e1001987.

56. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.

57. Hartling L, Bond K, Vandermeer B, et al. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. *PLoS One* 2011;6(2):e17242.

58. Hartling L, Hamm M, Milne A, et al. AHRQ Methods for Effective Health Care. Validity and Inter-Rater Reliability Testing of Quality Assessment Instruments. Rockville (MD): Agency for Healthcare Research and Quality (US) 2012.

59. Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol* 2013;66(9):973-81.

60. Jordan VM, Lensen SF, Farquhar CM. There were large discrepancies in risk of bias tool judgments when a randomized controlled trial appeared in more than one systematic review. *J Clin Epidemiol* 2017;81:72-76.

61. Kumar A, Miladinovic B, Guyatt GH, et al. GRADE guidelines system is reproducible when instructions are clearly operationalized even among the guidelines panel members with limited experience with GRADE. *J Clin Epidemiol* 2016;75:115-8.

62. Llewellyn A, Whittington C, Stewart G, et al. The Use of Bayesian Networks to Assess the Quality of Evidence from Research Synthesis: 2. Inter-Rater Reliability and Comparison with Standard GRADE Assessment. *PLoS One* 2015;10(12):e0123511.

63. Mustafa RA, Santesso N, Brozek J, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol* 2013;66(7):736-42; quiz 42.e1-5.

64. Norris SL, Holmer HK, Ogden LA, et al. AHRQ Methods for Effective Health Care. Selective Outcome Reporting as a Source of Bias in Reviews of Comparative Effectiveness. Rockville (MD): Agency for Healthcare Research and Quality (US) 2012.

65. O'Connor SR, Tully MA, Ryan B, et al. Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. *BMC Res Notes* 2015;8:224.

66. Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. *BMJ* 2013;346:f1798.

67. Whiting PF, Rutjes AW, Westwood ME, et al. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol* 2013;66(10):1093-104.

68. Sterne JAC, Egger M, Moher D, et al. Chapter 10: Addressing reporting biases. In: Higgins JPT, Churchill R, Chandler J, et al., eds. Cochrane Handbook for Systematic Reviews of Interventions version 5.2.0. (updated June 2017). Available from www.training.cochrane.org/handbook: Cochrane 2017.

69. Atakpo P, Vassar M. Publication bias in dermatology systematic reviews and meta-analyses. *J Dermatol Sci* 2016;82(2):69-74.

70. Lundh A, Gotzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. *BMC Med Res Methodol* 2008;8:22.

Figure legends

Figure 1. Flow diagram of identification, screening and inclusion of studies. [a]Refers to records identified from Ovid MEDLINE, Ovid EMBASE, Ovid PsycINFO, and Google Scholar. [b]Refers to records identified from screening references of included articles.

Table S1. Search strategies

Database: Ovid MEDLINE(R) <1946 to 9 February 2017>
Search Strategy:

----

1 ((tool or tools or instrument$ or checklist$ or check list$ or scale or scales) and (quality or methodolog$ or method or methods)).ti.
2 (quality adj10 (score or scores or scoring or rating or rate) adj5 (methodolog$ or method or methods)).tw.
3 (guideline$ and (quality or methodolog$ or method or methods)).ti.
4 ((assess$ or apprais$ or critical$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).ti.
5 ((score or scores or scoring or rating or rate) and (quality or methodolog$ or method or methods)).ti.
6 ((quality or methodology) adj3 (review or meta-analys$ or metaanalys$) adj3 (assess$ or method$)).tw.
7 (quality adj3 article$).tw.
8 (critical$ adj2 (apprais$ or evaluat$)).tw.
9 ((apprais$ or evaluat$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
10 (guideline$ adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
11 or/1-10
12 Checklist/
13 11 or 12
14 Publication Bias/
15 exp "bias (epidemiology)"/
16 (bias adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
17 ((quality or bias or methodolog$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
18 (bias$ adj3 (publication$ or disseminat$ or language$ or reporting or grey or gray or citation$ or time delay or time lag or conference or abstract)).tw.
19 or/14-18
20 13 and 19

Database: Embase <1980 to 2017 Week 06>
Search Strategy:

----

1 "Review Literature as Topic"/
2 "meta analysis (topic)"/
3 meta analysis/
4 "systematic review (topic)"/
5 systematic review/
6 systematic review$.tw.
7 (meta-analys$ or metaanalys$).tw.
8 or/1-7
9 (bias adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
10 ((quality or bias or methodolog$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
11 (bias$ adj3 (publication$ or disseminat$ or language$ or reporting or grey or gray or citation$ or time delay or time lag or conference or abstract)).tw.

12 "internal validity"/
13 publishing/
14 or/9-13
15 ((tool or tools or instrument$ or checklist$ or check list$ or scale or scales) and (quality or methodolog$ or method or methods)).ti.
16 (quality adj10 (score or scores or scoring or rating or rate) adj5 (methodolog$ or method or methods)).tw.
17 (guideline$ and (quality or methodolog$ or method or methods)).ti.
18 ((assess$ or apprais$ or critical$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).ti.
19 ((score or scores or scoring or rating or rate) and (quality or methodolog$ or method or methods)).ti.
20 ((quality or methodology) adj3 (review or meta-analys$ or metaanalys$) adj3 (assess$ or method$)).tw.
21 (quality adj3 article$).tw.
22 (critical$ adj2 (apprais$ or evaluat$)).tw.
23 ((apprais$ or evaluat$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
24 (guideline$ adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
25 or/15-24
26 checklist/
27 25 or 26
28 8 and 14 and 27
29 limit 28 to embase

Database: PsycINFO <1806 to February Week 1 2017>
Search Strategy:
----
1 meta-analysis/
2 systematic review$.tw.
3 (meta-analys$ or metaanalys$).tw.
4 or/1-3
5 ((tool or tools or instrument$ or checklist$ or check list$ or scale or scales) and (quality or methodolog$ or method or methods)).ti.
6 (quality adj10 (score or scores or scoring or rating or rate) adj5 (methodolog$ or method or methods)).tw.
7 (guideline$ and (quality or methodolog$ or method or methods)).ti.
8 ((assess$ or apprais$ or critical$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).ti.
9 ((score or scores or scoring or rating or rate) and (quality or methodolog$ or method or
1. methods)).ti.
10 ((quality or methodology) adj3 (review or meta-analys$ or metaanalys$) adj3 (assess$ or method$)).tw.
11 (quality adj3 article$).tw.
12 (critical$ adj2 (apprais$ or evaluat$)).tw.
13 ((apprais$ or evaluat$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
14 (guideline$ adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
15 checklist/
16 or/5-15
17 (bias adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.

18	((quality or bias or methodolog$) adj3 (systematic review$ or meta-analys$ or metaanalys$)).tw.
19	(bias$ adj3 (publication$ or disseminat$ or language$ or reporting or grey or gray or citation$ or time delay or time lag or conference or abstract)).tw.
20	bias.mp.
21	or/17-20
22	4 and 16 and 21

Table S2. Excluded studies

Reference	Reason for exclusion
Armijo-Olivo S, Cummings GG, Fuentes J, Saltaji H, Ha C, Chisholm A, et al. Identifying items to assess methodological quality in physical therapy trials: a factor analysis. Physical Therapy 2014;94(9):1272-84.	Paper does not report on a structured tool
Armijo-Olivo S, Fuentes J, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: a descriptive analysis. BMC Medical Research Methodology 2013;13:116.	Systematic review of tools
Armijo-Olivo S, Fuentes J, Rogers T, Hartling L, Saltaji H, Cummings GG. How should we evaluate the risk of bias of physical therapy trials?: a psychometric and meta-epidemiological approach towards developing guidelines for the design, conduct, and reporting of RCTs in Physical Therapy (PT) area: a study protocol. Syst Rev 2013;2:88.	Protocol for development of new tool
Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. International Journal of Evidence-Based Healthcare 2015;13(3):132-40.	Refers to a tool to assess quality of published systematic reviews
Arrive L, Renard R, Carrat F, Belkacem A, Dahan H, Le Hir P, et al. A scale of methodological quality for clinical studies of radiologic examinations. Radiology 2000;217(1):69-74.	Tool does not assess reporting bias
Atakpo P, Vassar M. Publication bias in dermatology systematic reviews and meta-analyses. Journal of Dermatological Science 2016;82(2):69-74.	Describes statistical methods only
Ballard M, Montgomery P. Risk of bias in overviews of reviews: a scoping review of methodological guidance and four-item checklist. Research Synthesis Methods 2017;8(1):92-108.	Refers to a tool to assess quality of published systematic reviews
Balzer K. Assessing the quality of research needs to go beyond scoring: Commentary on Crowe and Sheppard (2011). International Journal of Nursing Studies 2012;49(8):1048-50.	Commentary
Bartlett WA, Braga F, Carobene A, Coskun A, Prusa R, Fernandez-Calle P, et al. A checklist for critical appraisal of studies of biological variation. Clinical Chemistry and Laboratory Medicine 2015;53(6):879-85.	Tool does not assess reporting bias
Bashir R, Dunn AG. Systematic review protocol assessing the processes for linking clinical trial registries and their published results. BMJ Open 2016;6(10):e013048.	Paper does not report on a structured tool
Beck NB, Becker RA, Boobis A, Fergusson D, Fowle JR, Goodman J, et al. Instruments for assessing risk of bias and other methodological criteria of animal studies: omission of well-established methods. Environmental Health Perspectives 2014;122(3):A66-7.	Commentary

Reference	Reason for exclusion
Berkman ND, Lohr KN, Morgan LC, Kuo T-M, Morton SC. Interrater reliability of grading strength of evidence varies with the complexity of the evidence in systematic reviews. Journal of Clinical Epidemiology 2013;66(10):1105-17.e1.	Tool does not assess reporting bias
Burda BU, Holmer HK, Norris SL. Limitations of A Measurement Tool to Assess Systematic Reviews (AMSTAR) and suggestions for improvement. Systematic Reviews 2016;5:58.	Refers to a tool to assess quality of published systematic reviews
Cartes-Velasquez RA, Manterola C, Aravena P, Moraga J. Reliability and validity of MINCIR scale for methodological quality in dental therapy research. Brazilian Oral Research 2014;28.	Tool does not assess reporting bias
Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Research Synthesis Methods 2012;3(2):161-76.	Describes statistical methods only
da Costa BR, Hilfiker R, Egger M. PEDro's bias: summary quality scores should not be used in meta-analysis. Journal of Clinical Epidemiology 2013;66(1):75-7.	Commentary
Dahm P. Raising the bar for systematic reviews with Assessment of Multiple Systematic Reviews (AMSTAR). BJU International 2017;119(2):193.	Refers to a tool to assess quality of published systematic reviews
Dalton DR, Aguinis H, Dalton CM, Bosco FA, Pierce CA. Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. Personnel Psychology 2012;65(2):221-49.	Paper does not report on a structured tool
David SP, Ware JJ, Chu IM, Loftus PD, Fusar-Poli P, Radua J, et al. Potential reporting bias in fMRI studies of the brain. PloS One 2013;8(7):e70104.	Paper does not report on a structured tool
Davino-Ramaya C, Krause LK, Robbins CW, Harris JS, Koster M, Chan W, et al. Transparency matters: Kaiser Permanente's National Guideline Program methodological processes. The Permanente Journal 2012;16(1):55-62.	Refers to a tool to assess quality of published systematic reviews
Dawson A, Raphael KG, Glaros A, Axelsson S, Arima T, Ernberg M, et al. Development of a quality-assessment tool for experimental bruxism studies: reliability and validity. Journal of Orofacial Pain 2013;27(2):111-22.	Tool does not assess reporting bias
Deshpande S, Misso K, Westwood M, Stirk L, De Kock S, Clayton D, et al. Not all cochrane reviews are good quality systematic reviews. Value in Health 2016;19(7):A371.	Refers to a tool to assess quality of published systematic reviews
Disher T, Benoit B, Johnston C, Campbell-Yeo M. Skin-to-skin contact for procedural pain in neonates: acceptability of novel systematic review synthesis methods and GRADEing of the evidence. Journal of Advanced Nursing 2017;73(2):504-19.	Paper does not report on a structured tool
Dreier M, Borutta B, Stahmeyer J, Krauth C, Walter U. Comparison of tools for assessing the methodological quality of primary and	Systematic review of tools

Reference	Reason for exclusion
secondary studies in health technology assessment reports in Germany. GMS Health Technology Assessment 2010;6.	
Dreyer N, Velentgas P, Duddy A, Westrich KD, Dubois RW. Grace checklist: Rating the strength of evidence for observational studies of comparative effectiveness. Value in Health 2012;15(4):A5.	Tool does not assess reporting bias
Dreyer NA, Velentgas P, Westrich K, Dubois R. The GRACE checklist for rating the quality of observational studies of comparative effectiveness: a tale of hope and caution. Journal of Managed Care & Specialty Pharmacy 2014;20(3):301-8.	Tool does not assess reporting bias
Dreyer NA, Velentgas P, Westrich K, Dubois RW. GRACE: A validated checklist for identifying robust observational studies of comparative effectiveness. Pharmacoepidemiol Drug Saf 2013;22:356.	Tool does not assess reporting bias
Dreyer NA, Velentgas P, Westrich KD, Dubois RW. There but for grace? a validated screening tool for quality observational studies of comparative effectiveness. Value in Health 2013;16(3):A21.	Tool does not assess reporting bias
Drucker AM, Fleming P, Chan A-W. Research Techniques Made Simple: Assessing Risk of Bias in Systematic Reviews. The Journal of Investigative Dermatology 2016;136(11):e109-e14.	Guidance on using existing tools
Dwan K, Altman DG, Clarke M, Gamble C, Higgins JP, Sterne JA, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. PLoS Med 2014;11(6):e1001666.	Paper does not report on a structured tool
Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. PLoS One 2013;8(7):e66844.	Paper does not report on a structured tool
Dwan K, Kirkham JJ, Williamson PR, Gamble C. Selective reporting of outcomes in randomised controlled trials in systematic reviews of cystic fibrosis. BMJ Open 2013;3(6).	Evaluation of use of tool in practice, but no measurement properties assessed
Fantony JJ, Gopalakrishna A, Noord MV, Inman BA. Reporting Bias Leading to Discordant Venous Thromboembolism Rates in the United States Versus Non-US Countries Following Radical Cystectomy: A Systematic Review and Meta-analysis. European Urology Focus 2016;2(2):189-96.	Paper does not report on a structured tool
Fitzgerald A, Coop C. Validation and modification of the Graphical Appraisal Tool for Epidemiology (GATE) for appraising systematic reviews in evidence-based guideline development. Health Outcomes Research in Medicine 2011;2(1):e51-e9.	Refers to a tool to assess quality of published systematic reviews
Frosi G, Riley RD, Williamson PR, Kirkham JJ. Multivariate meta-analysis helps examine the impact of outcome reporting bias in Cochrane rheumatoid arthritis reviews. J Clin Epidemiol 2015;68(5):542-50.	Evaluation of use of tool in practice, but no measurement properties assessed
Furukawa TA, Miura T, Chaimani A, Leucht S, Cipriani A, Noma H, et al. Using the contribution matrix to evaluate complex study	Describes statistical methods only

Reference	Reason for exclusion
limitations in a network meta-analysis: a case study of bipolar maintenance pharmacotherapy review. BMC Res Notes 2016;9:218.	
Ghogomu EAT, Maxwell LJ, Buchbinder R, Rader T, Pardo Pardo J, Johnston RV, et al. Updated method guidelines for cochrane musculoskeletal group systematic reviews and metaanalyses. The Journal of Rheumatology 2014;41(2):194-205.	Guidance on using existing tools
Golder S, Loke YK, Bland M. Unpublished data can be of value in systematic reviews of adverse effects: methodological overview. Journal of Clinical Epidemiology 2010;63(10):1071-81.	Paper does not report on a structured tool
Golder S, Loke YK. Is there evidence for biased reporting of published adverse effects data in pharmaceutical industry-funded studies? British Journal of Clinical Pharmacology 2008;66(6):767-73.	Paper does not report on a structured tool
Goodyear-Smith FA, van Driel ML, Arroll B, Del Mar C. Analysis of decisions made in meta-analyses of depression screening and the risk of confirmation bias: a case study. BMC Med Res Methodol 2012;12:76.	Paper does not report on a structured tool
Grant S, Pedersen ER, Osilla KC, Kulesza M, D'Amico EJ. It is time to develop appropriate tools for assessing minimal clinically important differences, performance bias and quality of evidence in reviews of behavioral interventions. Addiction 2016;111(9):1533-5.	Paper does not report on a structured tool
Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostatistics (Oxford, England) 2001;2(4):463-71.	Describes statistical methods only
Haddaway NR, Woodcock P, Macura B, Collins A. Making literature reviews more reliable through application of lessons from systematic reviews. Conservation Biology 2015;29(6):1596-605.	Guidance on using existing tools
Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. Statistics in Medicine 2000;19(24):3325-36.	Describes statistical methods only
Heck NC, Mirabito LA, LeMaire K, Livingston NA, Flentje A. Omitted data in randomized controlled trials for anxiety and depression: A systematic review of the inclusion of sexual orientation and gender identity. Journal of Consulting and Clinical Psychology 2017;85(1):72-6.	Paper does not report on a structured tool
Higgins JPT, Lane PW, Anagnostelis B, Anzures-Cabrera J, Baker NF, Cappelleri JC, et al. A tool to assess the quality of a meta-analysis. Research Synthesis Methods 2013;4(4):351-66.	Refers to a tool to assess quality of published systematic reviews
Hoy D, Brooks P, Woolf A, Blyth F, March L, Bain C, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. J Clin Epidemiol 2012;65(9):934-9.	Tool does not assess reporting bias
Hsu W, Speier W, Taira RK. Automated extraction of reported statistical analyses: towards a logical representation of clinical trial	Paper does not report on a structured tool

Reference	Reason for exclusion
literature. AMIA Annual Symposium proceedings AMIA Symposium 2012;2012:350-9.	
Ioannidis JPA, Munafo MR, Fusar-Poli P, Nosek BA, David SP. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. Trends in Cognitive Sciences 2014;18(5):235-41.	Paper does not report on a structured tool
Ioannidis JPA, Trikalinos TA. An exploratory test for an excess of significant findings. Clinical Trials 2007;4(3):245-53.	Describes statistical methods only
Ioannidis JPA, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. CMAJ 2007;176(8):1091-6.	Describes statistical methods only
Jarde A, Losilla J-M, Vives J, Rodrigo MF. Q-Coh: A tool to screen the methodological quality of cohort studies in systematic reviews and meta-analysis. International Journal of Clinical and Health Psychology 2013;13(2):138-46.	Tool does not assess reporting bias
Jefferson T, Jones MA, Doshi P, Del Mar CB, Hama R, Thompson MJ, et al. Risk of bias in industry-funded oseltamivir trials: comparison of core reports versus full clinical study reports. BMJ Open 2014;4(9):e005253.	Evaluation of use of tool in practice, but no measurement properties assessed
Johnson BT, Low RE, MacDonald HV. Panning for the gold in health research: incorporating studies' methodological quality in meta-analysis. Psychology & Health 2015;30(1):135-52.	Describes statistical methods only
Johnston BC, Patrick DL, Busse JW, Schunemann HJ, Agarwal A, Guyatt GH. Patient-reported outcomes in meta-analyses--Part 1: assessing risk of bias and combining outcomes. Health and Quality of Life Outcomes 2013;11:109.	Guidance on using existing tools
Jorgensen L, Paludan-Muller AS, Laursen DR, Savovic J, Boutron I, Sterne JA, et al. Evaluation of the Cochrane tool for assessing risk of bias in randomized clinical trials: overview of published comments and analysis of user practice in Cochrane and non-Cochrane reviews. Syst Rev 2016;5:80.	Evaluation of use of tool in practice, but no measurement properties assessed
Jurgens T, Whelan AM, MacDonald M, Lord L. Development and evaluation of an instrument for the critical appraisal of randomized controlled trials of natural products. BMC Complement Altern Med 2009;9:11.	Tool does not assess reporting bias
Jurgens TM, Whelan AM. Development and evaluation of an instrument for the critical appraisal of randomized controlled trials of natural products. Canadian Journal of Hospital Pharmacy 2011;64(1):68.	Tool does not assess reporting bias
Katikireddi SV, Egan M, Petticrew M. How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. Journal of Epidemiology and Community Health 2015;69(2):189-95.	Audit of tools used in systematic reviews

Reference	Reason for exclusion
Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar S, Grimmer KA. A systematic review of the content of critical appraisal tools. BMC Med Res Methodol 2004;4:22.	Systematic review of tools
Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. Statistics in Medicine 2012;31(20):2179-95.	Describes statistical methods only
Kocsis JH, Gerber AJ, Milrod B, Roose SP, Barber J, Thase ME, et al. A new scale for assessing the quality of randomized clinical trials of psychotherapy. Comprehensive Psychiatry 2010;51(3):319-24.	Tool does not assess reporting bias
Kovacs FM, Abraira V. Language Bias in a Systematic Review of Chronic Pain: How to Prevent the Omission of Non-English Publications? The Clinical Journal of Pain 2004;20(3):199-200.	Paper does not report on a structured tool
Krauth D, Woodruff TJ, Bero L. Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. Environmental Health Perspectives 2013;121(9):985-92.	Systematic review of tools
Kromrey JD, Rendina-Gobioff G. On Knowing What We Do Not Know: An Empirical Comparison of Methods to Detect Publication Bias in Meta-Analysis. Educational and Psychological Measurement 2006;66(3):357-73.	Describes statistical methods only
Lamont RF. A quality assessment tool to evaluate tocolytic studies. BJOG 2006;113(Suppl 3):96-9.	Tool does not assess reporting bias
Langendam M, Carrasco-Labra A, Santesso N, Mustafa RA, Brignardello-Petersen R, Ventresca M, et al. Improving GRADE evidence tables part 2: A systematic survey of explanatory notes shows more guidance is needed. J Clin Epidemiol 2016;74:19-27.	Evaluation of use of tool in practice, but no measurement properties assessed
Liebherz S, Schmidt N, Rabung S. How to assess the quality of psychotherapy outcome studies: A systematic review of quality assessment criteria. Psychotherapy Research 2016;26(5):573-89.	Systematic review of tools
Liebherz S, Schmidt N, Rabung S. Study Quality and its Influence on Treatment Outcome in Studies on the Effectiveness of Inpatient Psychotherapy - A Meta-Analysis. PPmP Psychotherapie Psychosomatik Medizinische Psychologie 2016;66(1):31-8.	Not written in English
Lohr KN, Carey TS. Assessing "best evidence": issues in grading the quality of studies for systematic reviews. The Joint Commission Journal on Quality Improvement 1999;25(9):470-9.	Guidance on using existing tools
Lonjon G, Porcher R, Ergina P, Fouet M, Boutron I. Potential Pitfalls of Reporting and Bias in Observational Studies With Propensity Score Analysis Assessing a Surgical Procedure: A Methodological Systematic Review. Ann Surg 2016:no pagination.	Paper does not report on a structured tool
Lundh A, Gotzsche PC. Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. BMC Med Res Methodol 2008;8:22.	Guidance on using existing tools

Reference	Reason for exclusion
Lynch HN, Goodman JE, Tabony JA, Rhomberg LR. Systematic comparison of study quality criteria. Regul Toxicol Pharmacol 2016;76:187-98.	Systematic review of tools
Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, et al. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. PLoS Biology 2015;13(10):e1002273.	Tool does not assess reporting bias
Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther 2003;83(8):713-21.	Tool does not assess reporting bias
Malmivaara A. Methodological considerations of the GRADE method. Annals of Medicine 2015;47(1):1-5.	Guidance on using existing tools
Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. Journal of the American Medical Informatics Association 2016;23(1):193-201.	Model to semi-automate Cochrane risk of bias tool
McDonagh MS, Peterson K, Balshem H, Helfand M. US Food and Drug Administration documents can provide unpublished evidence relevant to systematic reviews. Journal of Clinical Epidemiology 2013;66(10):1071-81.	Paper does not report on a structured tool
McShane BB, Bockenholt U, Hansen KT. Adjusting for Publication Bias in Meta-Analysis: An Evaluation of Selection Methods and Some Cautionary Notes. Perspectives on Psychological Science 2016;11(5):730-49.	Describes statistical methods only
Millard LAC, Flach PA, Higgins JPT. Machine learning to assist risk-of-bias assessments in systematic reviews. International Journal of Epidemiology 2016;45(1):266-77.	Model to semi-automate Cochrane risk of bias tool
Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Controlled Clinical Trials 1995;16(1):62-73.	Systematic review of tools
Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLoS Med 2014;11(10):e1001744.	Refers to a tool to assess quality of published systematic reviews
Moyer A, Finney JW. Rating methodological quality: toward improved assessment and investigation. Accountability in Research 2005;12(4):299-313.	Guidance on using existing tools
Mueller KF, Briel M, Strech D, Meerpohl JJ, Lang B, Motschall E, et al. Dissemination bias in systematic reviews of animal research: a systematic review. PloS One 2014;9(12):e116016.	Paper does not report on a structured tool
Mueller KF, Meerpohl JJ, Briel M, Antes G, von Elm E, Lang B, et al. Detecting, quantifying and adjusting for publication bias in meta-analyses: protocol of a systematic review on methods. Systematic Reviews 2013;2:60.	Describes statistical methods only

Reference	Reason for exclusion
Mueller KF, Meerpohl JJ, Briel M, Antes G, von Elm E, Lang B, et al. Methods for detecting, quantifying, and adjusting for dissemination bias in meta-analysis are described. J Clin Epidemiol 2016;80:25-33.	Describes statistical methods only
Nakagawa S, Noble DWA, Senior AM, Lagisz M. Meta-evaluation of meta-analysis: ten appraisal questions for biologists. BMC Biology 2017;15(1):18.	Refers to a tool to assess quality of published systematic reviews
Nolting A, Perleth M, Langer G, Meerpohl JJ, Gartlehner G, Kaminski-Hartenthaler A, et al. [GRADE guidelines: 5. Rating the quality of evidence: publication bias]. Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen 2012;106(9):670-6.	Not written in English
Norris SL, Moher D, Reeves BC, Shea B, Loke Y, Garner S, et al. Issues relating to selective reporting when including non-randomized studies in systematic reviews on the effects of healthcare interventions. Res Synth Methods 2013;4(1):36-47.	Guidance on using existing tools
Nurmatov UB, Xiong T, Kroes MA. Evaluation of quality assessment tools for non-randomised controlled trials assessing surgical interventions: A systematic review of systematic reviews. Value in Health 2015;18(7):A722.	Systematic review of tools
Odierna DH, Forsyth SR, White J, Bero LA. The cycle of bias in health research: a framework and toolbox for critical appraisal training. Accountability in Research 2013;20(2):127-41.	Paper does not report on a structured tool
Palma Perez S, Delgado Rodriguez M. [Practical considerations on detection of publication bias]. Gac Sanit 2006;20(Suppl 3):10-6.	Not written in English
Pearson M, Peters J. Outcome reporting bias in evaluations of public health interventions: evidence of impact and the potential role of a study register. Journal of Epidemiology and Community Health 2012;66(4):286-9.	Evaluation of use of tool in practice, but no measurement properties assessed
Petticrew M, Egan M, Thomson H, Hamilton V, Kunkler R, Roberts H. Publication bias in qualitative research: what becomes of qualitative research presented at conferences? Journal of Epidemiology and Community Health 2008;62(6):552-4.	Paper does not report on a structured tool
Pigott TD, Valentine JC, Polanin JR, Williams RT, Canada DD. Outcome-Reporting Bias in Education Research. Educational Researcher 2013;42(8):424-32.	Paper does not report on a structured tool
Pirracchio R, Resche-Rigon M, Chevret S, Journois D. Do simple screening statistical tools help to detect reporting bias? Annals of Intensive Care 2013;3(1):29.	Describes statistical methods only
Quigley JM, Thompson J, Halfpenny N, Scott DA. Critical appraisal of non-randomized controlled trials-a review of recommended and commonly used tools. Value in Health 2014;17(3):A203.	Systematic review of tools
Quigley JM, Thompson JC, Halfpenny NJ, Scott DA. Critical appraisal of real world evidence-a review of recommended and commonly used tools. Value in Health 2015;18(7):A684.	Systematic review of tools

Reference	Reason for exclusion
Quintana DS. From pre-registration to publication: A non-technical primer for conducting a meta-analysis to synthesize correlational data. Front Psychol 2015;6:1549.	Paper does not report on a structured tool
Rangel SJ, Kelsey J, Colby CE, Anderson J, Moss RL. Development of a quality assessment scale for retrospective clinical studies in pediatric surgery. Journal of Pediatric Surgery 2003;38(3):390-6.	Tool does not assess reporting bias
Rosella L, Bowman C, Pach B, Morgan S, Fitzpatrick T, Goel V. The development and validation of a meta-tool for quality appraisal of public health evidence: Meta Quality Appraisal Tool (MetaQAT). Public Health 2016 Jul;136:57-65.	Tool does not assess reporting bias
Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. International Journal of Epidemiology 2007;36(3):666-76.	Systematic review of tools
Santaguida PL, Riley CM, Matchar DB. Chapter 5: Assessing risk of bias as a domain of quality in medical test studies. Journal of General Internal Medicine 2012;27(Suppl 1):S33-S8.	Guidance on using existing tools
Savovic J, Weeks L, Sterne JA, Turner L, Altman DG, Moher D, et al. Evaluation of the Cochrane Collaboration's tool for assessing the risk of bias in randomized trials: focus groups, online survey, proposed recommendations and their implementation. Syst Rev 2014;3:37.	Evaluation of use of tool in practice, but no measurement properties assessed
Seehra J, Pandis N, Koletsi D, Fleming PS. Use of quality assessment tools in systematic reviews was varied and inconsistent. J Clin Epidemiol 2016;69:179-84.e5.	Audit of tools used in systematic reviews
Shamliyan T, Kane RL, Dickinson S. A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. Journal of Clinical Epidemiology 2010;63(10):1061-70.	Systematic review of tools
Shamliyan TA, Kane RL, Ansari MT, Raman G, Berkman ND, Grant M, et al. Development quality criteria to evaluate nontherapeutic studies of incidence, prevalence, or risk factors of chronic diseases: pilot study of new checklists. Journal of Clinical Epidemiology 2011;64(6):637-57.	Tool does not assess reporting bias
Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10.	Refers to a tool to assess quality of published systematic reviews
Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. Journal of Clinical Epidemiology 2009;62(10):1013-20.	Refers to a tool to assess quality of published systematic reviews
Shuang M, Zhao C, Zhang L, Shang HC. Using SYRCLE tools to evaluate the methodological quality of animal experiments of stroke in China. Chinese Journal of Evidence-Based Medicine 2016;16(5):592-7.	Not written in English

Reference	Reason for exclusion
Singh S, Khosla S. Suboptimal choice of methodology for meta-analysis and publication bias assessment. The American Journal of Cardiology 2015;115(12):1782-3.	Describes statistical methods only
Smyth RM, Kirkham JJ, Jacoby A, Altman DG, Gamble C, Williamson PR. Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists. BMJ 2011;342:c7153.	Paper does not report on a structured tool
Sohani ZN, Meyre D, de Souza RJ, Joseph PG, Gandhi M, Dennis BB, et al. Assessing the quality of published genetic association studies in meta-analyses: the quality of genetic studies (Q-Genie) tool. BMC Genet 2015;16:50.	Tool does not assess reporting bias
Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technology Assessment (Winchester, England) 2010;14(8):iii-193.	Paper does not report on a structured tool
Spooner CH, Pickard AS, Menon D. Edmonton Quality Assessment Tool for Drug Utilization Reviews: EQUATDUR-2: the development of a scale to assess the methodological quality of a drug utilization review. Medical Care 2000;38(9):948-58.	Tool does not assess reporting bias
Tate RL, Perdices M, Rosenkoetter U, Wakim D, Godbee K, Togher L, et al. Revision of a method quality rating scale for single-case experimental designs and n-of-1 trials: the 15-item Risk of Bias in N-of-1 Trials (RoBiNT) Scale. Neuropsychological Rehabilitation 2013;23(5):619-38.	Tool does not assess reporting bias
Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters M, et al. AHRQ Methods for Effective Health Care Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012.	Guidance on using existing tools
Voss PH, Rehfuess EA. Quality appraisal in systematic reviews of public health interventions: an empirical study on the impact of choice of tool on meta-analysis. Journal of Epidemiology and Community Health 2013;67(1):98-104.	Evaluation of existing tools
Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2008. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed 7/03/2017).	Tool does not assess reporting bias
Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J. A systematic review finds that diagnostic reviews fail to incorporate quality despite available tools. Journal of Clinical Epidemiology 2005;58(1):1-12.	Systematic review of tools
Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of	Tool does not assess reporting bias

Reference	Reason for exclusion
studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.	
Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of Internal Medicine 2011;155(8):529-36.	Tool does not assess reporting bias
Wiart L, Kolaski K, Vogtle LK, Butler C, Romeiser Logan L, Hickman R, et al. Inter-rater reliability and concurrent validity of the AACPDM study design and quality rating system for conducting systematic reviews (group design). Dev Med Child Neurol 2011;53:74.	Refers to a tool to assess quality of published systematic reviews

Table S3. General characteristics of included tools

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
Balshem 2013[1]	AHRQ outcome and analysis reporting bias framework	Domain-based	Reporting bias only	Bias due to selective non-reporting and bias in selection of the reported result	Randomized trials	Specific outcome/ result in a study	Expert consensus (via email)	Brief annotation per item/response option	No
Berkman 2013[2]	AHRQ tool for evaluating the risk of reporting bias	Domain-based	Reporting bias only	Bias due to selective publication and bias due to selective non-reporting	Systematic reviews	Specific synthesis of studies	Not stated	Brief annotation per item/response option	No
Downes 2016[3]	AXIS tool (Appraisal tool for Cross-Sectional Studies)	Checklist	Multiple sources of bias	Bias due to selective non-reporting	Cross-sectional studies	Whole study	Literature review, piloting, Delphi study	None	No
Downs 1998[4]	Downs-Black tool	Scale	Multiple sources of bias	Bias in selection of the	Randomized trials and non-	Whole study	Literature review, piloting,	Brief annotation per	Yes

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
				reported result	randomized studies of interventions		psychometric testing	item/response option	
Guyatt 2011[5-9]	GRADE	Domain-based	Multiple sources of bias	Bias due to selective publication and bias due to selective non-reporting	Systematic reviews	Specific synthesis of studies	Literature review, expert consensus (face-to-face and email), user testing	Detailed guidance manual	Yes
Hayden 2013[10]	QUIPS (Quality In Prognosis Studies) tool	Domain-based	Multiple sources of bias	Bias due to selective non-reporting	Prognosis studies	Whole study	Modified Delphi approach, nominal group technique at facilitated discussion workshop; piloting	Brief annotation per item/response option	Yes
Higgins 2008[11-13]	Cochrane risk of bias tool for randomized trials	Domain-based	Multiple sources of bias	Bias due to selective non-reporting and bias in selection of the	Randomized trials	Whole study	Literature review, informal consensus at facilitated meeting, piloting, focus groups and	Detailed guidance manual	Yes

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
				reported result			surveys, followed by consensus meeting		
Higgins 2016[14 15]	RoB 2.0 (revised tool for assessing risk of bias in randomized trials)	Domain-based	Multiple sources of bias	Bias in selection of the reported result	Randomized trials	Specific outcome/ result in a study	Literature review, informal consensus at facilitated meeting, piloting	Detailed guidance manual	No
Hoojimans 2014[16]	SYRCLE's RoB tool (SYstematic Review Centre for Laboratory animal Experimentation)	Domain-based	Multiple sources of bias	Bias due to selective non-reporting and bias in selection of the reported result	Animal studies	Whole study	Adaptation of existing tool, literature review	Brief annotation per item/response option	No
Kim 2013[17]	RoBANS (Risk of Bias Assessment Tool for Nonrandomized Studies)	Domain-based	Multiple sources of bias	Bias due to selective non-reporting and bias in selection of the	Non-randomized studies of interventions	Whole study	Literature review, psychometric testing	Brief annotation per item/response option	Yes

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
				reported result					
Kirkham 2010[18 19]	ORBIT-I (Outcome Reporting Bias In Trials) classification system for benefit outcomes	Domain-based	Reporting bias only	Bias due to selective non-reporting	Randomized trials	Specific outcome/ result in a study	Iteratively developed as part of a methodological study	Worked example for each response option	Yes
Meader 2014[20 21]	SAQAT (Semi-Automated Quality Assessment Tool)	Domain-based	Multiple sources of bias	Bias due to selective publication and bias due to selective non-reporting	Systematic reviews	Specific synthesis of studies	Development of logic model based on GRADE articles and piloting	None	Yes
Reid 2015[22]	Selective reporting bias algorithm	Domain-based	Reporting bias only	Bias due to selective non-reporting and bias in selection of the	Randomized trials	Whole study	Not stated	Brief annotation per item/response option	No

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
				reported result					
Saini 2014[23]	ORBIT-II (Outcome Reporting Bias In Trials) classification system for harm outcomes	Domain-based	Reporting bias only	Bias due to selective non-reporting	Randomized trials and non-randomized studies of interventions	Specific outcome/ result in a study	Iteratively developed as part of a methodological study	Worked example for each response option	No
Salanti 2014[24 25]	Framework for evaluating the quality of evidence from a network meta-analysis	Domain-based	Multiple sources of bias	Bias due to selective publication and bias due to selective non-reporting	Network meta-analyses	Specific synthesis of studies	Adaptation of existing tool	Detailed annotation per item/response option	No
Sterne 2016[26]	ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool	Domain-based	Multiple sources of bias	Bias in selection of the reported result	Non-randomized studies of interventions	Specific outcome/ result in a study	Expert consensus meetings (face-to-face), piloting	Detailed guidance manual	Yes

Article ID	Tool	Type of tool	Scope of tool	Types of reporting bias	Types of study designs	Level of assessment	Methods used to develop tool	Guidance available	Measurement properties evaluated
Viswanathan 2012[27]	RTI Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures	Domain-based	Multiple sources of bias	Bias due to selective non-reporting	Non-randomized studies of interventions or exposures	Whole study	Literature review, expert consensus (via email), cognitive testing, psychometric testing	Brief annotation per item/response option	No
Viswanathan 2013[28]	RTI Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures	Domain-based	Multiple sources of bias	Bias due to selective non-reporting	Non-randomized studies of interventions or exposures	Whole study	Literature review, expert consensus (via email)	Brief annotation per item/response option	No

References

1. Balshem H, Stevens A, Ansari M, et al. Finding grey literature evidence and assessing for outcome and analysis reporting biases when comparing medical interventions: AHRQ and the Effective Health Care Program. (Prepared by the Oregon Health and Science University and the University of Ottawa Evidence-based Practice Centers under Contract Nos. 290-2007-10057-I and 290-2007-10059-I.) AHRQ Publication No. 13(14)-EHC096-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

2. Berkman ND, Lohr KN, Ansari M, et al. Chapter 15 Appendix A: A Tool for Evaluating the Risk of Reporting Bias (in Chapter 15: Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update). Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I). AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm

3. Downes MJ, Brennan ML, Williams HC, et al. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ open* 2016;6:e011458.

4. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.

5. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.

6. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64(12):1277-82.

7. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol* 2011;64(4):407-15.

8. Schünemann H, Brożek J, Guyatt G, et al. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. [Updated October 2013]. Available from http://gdt.guidelinedevelopment.org/app/handbook/handbook.html.

9. Santesso N, Carrasco-Labra A, Langendam M, et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. *J Clin Epidemiol* 2016

10. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.

11. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons 2008:187-241.

12. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org/.

13. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

14. Higgins JPT, Savović J, Page MJ, et al. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0), Version 20 October 2016. Available from http://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/rob2-0/ [accessed 19 September 2017].

15. Higgins JPT, Sterne JAC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials. *Cochrane Methods Cochrane Database of Systematic Reviews* 2016;10(Suppl 1):29-31.

16. Hooijmans CR, Rovers MM, de Vries RBM, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.

17. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013;66(4):408-14.

18. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.

19. Dwan K, Gamble C, Kolamunnage-Dona R, et al. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials* 2010;11:52.

20. Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic reviews* 2014;3(1):82.

21. Stewart GB, Higgins JP, Schunemann H, et al. The use of Bayesian networks to assess the quality of evidence from research synthesis: 1. *PLoS One* 2015;10(3):e0114497.

22. Reid EK, Tejani AM, Huan LN, et al. Managing the incidence of selective reporting bias: a survey of Cochrane review groups. *Systematic reviews* 2015;4:85.

23. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ* 2014;349:g6501.

24. Salanti G, Giovane CD, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9(7):e99682.

25. Higgins JP, Del Giovane C, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *Value Health* 2014;17(7):A324.

26. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.

27. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* 2012;65(2):163-78.

28. Viswanathan M, Berkman ND, Dryden DM, et al. AHRQ Methods for Effective Health Care. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or

Exposures: Further Development of the RTI Item Bank. Rockville (MD): Agency for Healthcare Research and Quality (US) 2013.

Table S4. Items and response options relating to risk of reporting biases

Article ID	Tool	Items	Response options
Balsham 2013[1]	AHRQ outcome and analysis reporting bias framework	1. Across all study source documents, what is the risk of ORB/ARB? Compare published report(s) against (1) study protocol (if not retrieved in literature search), (2) trial registry entry/regulatory documents/industry documents, (3) other sources if applicable. 2. If ORB risk unclear: Given the study objectives, duration, and other investigated outcomes, could the study have also likely measured the outcome of interest but not reported it?	**Outcome reporting bias risk positive (ORB risk +)**: If reviewers determine that an outcome X was planned but the results were not reported, or were only partially reported in study documents, then the study is at risk of reporting bias for that outcome ("ORB risk +"). Also, if reviewers determine that an outcome X was not planned but the results were reported, then the study is at risk of reporting bias for that outcome ("ORB risk +"). Also, for studies for which the risk of reporting bias cannot be ruled out, reviewers should ask the question: "Given the study objectives, duration, and other investigated outcomes, could the study have also likely measured the outcome of interest but not reported it?" When the answer is "yes" (e.g. another reported outcome in the study leads the reviewer to believe that outcome X would have been collected), then the study should be rated "ORB risk +" for that outcome. **Outcome reporting bias risk negative (ORB risk -)**: When it is clear to the reviewers that outcome X was planned (e.g. from protocol, regulatory submissions, etc.), complete outcome data are available from at least one study document (published or otherwise), and the outcome was appropriately analyzed as planned, then the study is not at risk for reporting bias for this outcome. Also, for studies for which the risk of reporting bias cannot be ruled out, reviewers should ask the question: "Given the study objectives, duration, and other investigated outcomes, could the study have also

Article ID	Tool	Items	Response options
			likely measured the outcome of interest but not reported it?" If the answer is "no" the study should be rated as "ORB risk– **Outcome reporting bias risk unclear (ORB risk unclear)**: If the reviewers are unable to determine whether an outcome X was planned, but data are reported completely or partially, then the study risk of outcome and analysis reporting bias may be categorized as "unclear". This would also apply to a study that did not report any outcome of review interest across all source documents but was eligible on population, intervention, comparator, and other criteria. Also, for studies for which the risk of reporting bias cannot be ruled out, reviewers should ask the question: "Given the study objectives, duration, and other investigated outcomes, could the study have also likely measured the outcome of interest but not reported it?" If it still remains unclear whether the outcome of interest may have been assessed, the study should be categorized as "ORB risk unclear." **Analysis reporting bias risk positive (ARB risk +)**: When reported results are based on a different analysis, effect measure, cut-off, etc. than what was prespecified, then the study is at risk of analysis reporting bias for that outcome ("ARB risk +"). A study is also at risk of analysis reporting ("ARB risk +") because there is no way to know whether the reported analysis was planned or post hoc. **Analysis reporting bias risk negative (ARB risk -)**: When it is clear to the reviewers that outcome X was planned (e.g. from protocol, regulatory submissions, etc.),

Article ID	Tool	Items	Response options
			complete outcome data are available from at least one study document (published or otherwise), and the outcome was appropriately analyzed as planned, then the study is not at risk for reporting bias for this outcome **Analysis reporting bias risk unclear (ARB risk unclear)**: If the reviewers are unable to determine whether an outcome X was planned, but data are reported completely or partially, then the study risk of outcome and analysis reporting bias may be categorized as "unclear". This would also apply to a study that did not report any outcome of review interest across all source documents but was eligible on population, intervention, comparator, and other criteria.
Berkman 2013[2]	AHRQ tool for evaluating the risk of reporting bias	1. Are all the following criteria met: ≥10 studies contributing data for an outcome, studies of unequal sizes, no substantial clinical and methodological differences between smaller and larger studies, and quantitative results accompanied with measures of dispersion? 2. If yes, do smaller studies tend to demonstrate more favorable results? (visual assessment) 3. If yes, what is the result of a test for funnel plot asymmetry? 4. If test is positive, would a clinical decision differ for estimates from a fixed effects versus random effect model because the findings from a fixed effect model are closer to the null? 5. If no to the first question, is there an explanation for substantial heterogeneity?	**Suspected risk of reporting bias**: Testing for funnel plot asymmetry demonstrates a substantial likelihood of bias, and/or a qualitative assessment suggests the likelihood of missing studies, analyses, or outcomes data that may alter the conclusions from the reported evidence. **Undetected risk of reporting bias**: All alternative scenarios.

Article ID	Tool	Items	Response options
		6. If no to any of Q1-5, what is the estimated N of studies that are affected by SOR, SAR, nonpublication, or nonaccessibility? 7. If no to any of Q1-5, what is the total sample size of evidence affected by reporting bias (when known)? 8. If no to any of Q1-5, what is the total N of studies in evidence base? 9. If no to any of Q1-5, what is the total N of participants in evidence base? 10. If no to any of Q1-5, what is the consistency of effect estimates across contributing studies? 11. If no to any of Q1-5, what are the study limitations for the evidence base? 12. If no to any of Q1-5, what is the comprehensiveness of study retrieval and identification?	
Downes 2016[3]	AXIS tool (Appraisal tool for Cross-Sectional Studies)	1. Were the results for the analyses described in the methods, presented?	**Yes**: Not stated **No**: Not stated **Do not know/comment**: Not stated
Downs 1998[4]	Downs-Black tool	1. If any of the results of the study were based on "data dredging", was this made clear?	**Yes**: Any analyses that had not been planned at the outset of the study were clearly indicated. Also, no retrospective unplanned subgroup analyses were reported. **No**: Any analyses that had not been planned at the outset of the study were not clearly indicated. **Unable to determine**: Not stated
Guyatt 2011[5-9]	GRADE	1. Study limitations (including selective outcome reporting) 2. Publication bias	**Study limitations domain – No serious limitations, do not downgrade**: Most information is from studies at low risk of bias (i.e. those with low risk of bias for all key

Article ID	Tool	Items	Response options
			criteria, including lack of allocation concealment, lack of blinding, incomplete accounting of patients and outcome events, selective outcome reporting bias, other limitations [stopping early for benefit, use of unvalidated outcome measures, carryover effects in crossover trial, recruitment bias in cluster-randomized trial]) **Study limitations domain – Serious limitations, rate down one level (i.e., from high to moderate quality)**: Most information is from studies at moderate risk of bias **Study limitations domain – Very serious limitations, rate down two levels (i.e., from high to low quality or moderate to very low)**: Most information is from studies at high risk of bias. Selective reporting is present if authors acknowledge prespecified outcomes that they fail to report or report outcomes incompletely such that they cannot be included in a metaanalysis. One should suspect reporting bias if the study report fails to include results for a key outcome that one would expect to see in such a study or if composite outcomes are presented without the individual component outcomes. **Publication bias domain – Undetected**: None of the criteria for “strongly suspected” are met **Publication bias domain – Strongly suspected**: “In general, review authors and guideline developers should consider rating down for likelihood of publication bias when the evidence consists of a number of small studies. The inclination to rate down for publication bias should increase if most of those small studies are industry sponsored or likely to be industry sponsored (or if the investigators share another conflict of interest)...Another

Article ID	Tool	Items	Response options
			criterion for publication bias is the pattern of study results. Suspicion may increase if visual inspection demonstrates an asymmetrical rather than a symmetrical funnel plot or if statistical tests of asymmetry are positive. Although funnel plots may be helpful, review authors and guideline developers should bear in mind that visual assessment of funnel plots is distressingly prone to error. Enhancements of funnel plots may (or may not) help to improve reproducibility and validity associated with their use...Furthermore, systematic review and guideline authors should bear in mind that even if they find convincing evidence of asymmetry, publication bias is not the only explanation. For instance, if smaller studies suffer from greater study limitations, they may yield biased overestimates of effects. Another explanation would be that, because of a more restrictive (and thus responsive) population, or a more careful administration of the intervention, the effect may actually be larger in the small studies...More compelling than any of these theoretical exercises is authors' success in obtaining the results of some unpublished studies and demonstrating that the published and unpublished data show different results. In these circumstances, the possibility of publication bias looms large. The risk of publication bias is probably larger for observational studies than for RCTs, particularly small observational studies and studies conducted on data collected automatically (e.g. in the electronic medical record or in a diabetes registry) or data collected for a previous study. In these instances, it is difficult for the reviewer to know if the observational studies that appear in the literature represent all or a

Article ID	Tool	Items	Response options
			fraction of the studies conducted, and whether the analyses in them represent all or a fraction of those conducted. In these instances, reviewers may consider the risk of publication bias as substantial” [6]. “Guideline panels and authors of systematic reviews should consider the extent to which they are uncertain about the magnitude of the effect due to selective publication of studies and they may downgrade the quality of evidence by one level. Consider: study design (experimental vs. observational); study size (small studies vs. large studies); lag bias (early publication of positive results); search strategy (was it comprehensive?); asymmetry in funnel plot” [8]. “Relevant content: whether publication bias is undetected or suspected; interpretation of funnel plot; comprehensiveness of the search strategies and methods to identify all available evidence; presence of small (often positive) studies with for profit interest...Indicate the reason publication bias is detected (e.g. asymmetrical funnel plot, small studies with positive results, suspected selective availability of data from published, or unpublished studies)” [9].
Hayden 2013[10]	QUIPS (Quality In Prognosis Studies) tool	1. Statistical analysis and reporting (the statistical analysis is appropriate and all primary outcomes are reported). Prompting items include (a) Sufficient presentation of data to assess the adequecy of the analytic strategy; (b) Strategy for model building is appropriate and is based on a conceptual framework or model; (c) The selected statistical model is adequate for the design of the study; (d) There is no selective reporting of results.	**Low risk of bias**: The reported results are unlikely to be spurious or biased related to analysis or reporting **Moderate risk of bias**: The reported results may be spurious or biased related to analysis or reporting **High risk of bias**: The reported results are very likely to be spurious or biased related to analysis or reporting

Article ID	Tool	Items	Response options
Higgins 2008[11-13]	Cochrane risk of bias tool for randomized trials	1. Are reports of the study free of suggestion of selective outcome reporting? (2008 version); Reporting bias due to selective outcome reporting (2011 version)	**Low risk of bias**: Any of the following – The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way; The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon). **High risk of bias**: Any one of the following – Not all of the study's pre-specified primary outcomes have been reported; One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not prespecified; One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; The study report fails to include results for a key outcome that would be expected to have been reported for such a study. **Unclear risk of bias**: Insufficient information to permit judgement of 'Low risk' or 'High risk'. It is likely that the majority of studies will fall into this category.
Higgins 2016[14 15]	RoB 2.0	1. Are the reported outcome data likely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain, or from multiple analyses of the data?	**Low risk of bias**: Reported outcome data are unlikely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain, and reported outcome data are unlikely to have been selected, on the basis of the results, from multiple analyses of the data.

Article ID	Tool	Items	Response options
			High risk of bias: Reported outcome data are likely to have been selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain, or from multiple analyses of the data (or both). **Some concerns**: There is insufficient information available to exclude the possibility that reported outcome data were selected, on the basis of the results, from multiple outcome measurements (e.g. scales, definitions, time points) within the outcome domain, or from multiple analyses of the data.
Hoojimans 2014[16]	SYRCLE's RoB tool (SYstematic Review Centre for Laboratory animal Experimentation)	1. Are reports of the study free of selective outcome reporting? Includes two signalling questions: Was the study protocol available and were all of the study's pre-specified primary and secondary outcomes reported in the current manuscript?; Was the study protocol not available, but was it clear that the published report included all expected outcomes (i.e. comparing methods and results section)?	**Low risk of bias**: Not stated, but assume same criteria as Cochrane risk of bias tool for randomized trials [13]. **High risk of bias**: Not all of the study's pre-specified primary outcomes have been reported; One or more primary outcomes have been reported using measurements, analysis methods or data subsets (e.g. subscales) that were not pre-specified in the protocol; One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting has been provided, such as an unexpected adverse effect); The study report fails to include results for a key outcome that would be expected to have been reported for such a study. **Unclear risk of bias**: Not stated, but assume same criteria as Cochrane risk of bias tool for randomized trials [13].
Kim 2013[17]	RoBANS (Risk of Bias Assessment	1. Reporting biases caused by the selective reporting of outcomes	**Low risk of bias**: Any one of the following conditions – The experimental protocol is available, and the pre-

Article ID	Tool	Items	Response options
	Tool for Nonrandomized Studies)		defined primary/secondary outcomes were described as planned; All of the expected outcomes were included in the study descriptions (even in the absence of the experimental protocols). **High risk of bias**: Any one of the following conditions – The pre-defined primary outcomes were not fully reported; The outcomes were not reported in accordance with the previously defined standards; Primary outcomes that were not pre-specified in the study existed (except for outcomes with clear explanations, such as unexpected adverse effects); The existence of incomplete reporting regarding the primary outcome of interest; The absence of reports on important outcomes that would be expected to be reported for studies in related fields. **Unclear risk of Bias**: It is uncertain whether the selective outcome reporting resulted in a 'high risk' or a 'low risk' of bias.
Kirkham 2010[18 19]	ORBIT-I (Outcome Reporting Bias In Trials) classification system for benefit outcomes	1. The Outcome Reporting Bias In Trials (ORBIT) study classification system for missing or incomplete outcome reporting in reports of randomised trials	**Low risk of bias**: A "low risk" classification was awarded when it was suspected, but not actually known, that the outcome was either not measured, measured but not analysed, or measured and analysed but either partially reported or not reported for a reason unrelated to the results obtained. Specific examples include: (C) Trial report states that outcome was analysed but insufficient data were presented for the trial to be included in meta-analysis or to be considered to be fully tabulated; (F) Clear that outcome was measured but not necessarily analysed, and judgment says unlikely to have been analysed but not reported because of non-significant

Article ID	Tool	Items	Response options
			results; (H) Not mentioned but clinical judgment says outcome unlikely to have been measured at all. **High risk of bias**: A "high risk" classification was awarded when it was either known or suspected that the results were partially or not reported because the treatment comparison was statistically non-significant (P>0.05). Specific examples include: (A) Trial report states that outcome was analysed but only reports that result was not significant (typically stating P>0.05); (D) Trial report states that outcome was analysed but no results reported; (E) Clear that outcome was measured but not necessarily analysed, and judgment says likely to have been analysed but not reported because of non-significant results; (G) Not mentioned but clinical judgment says outcome likely to have been measured and analysed but not reported on the basis of non-significant results. **No risk of bias**: A "no risk" classification was reserved for cases where it was known that the outcome was not measured, known that it was measured but not analysed, or known that it was measured and analysed but the reason for partial or no reporting was not because the results were statistically non-significant. Specific examples include: (B) Trial report states that outcome was analysed but only reports that result was significant (typically stating P<0.05); (I) Clear that outcome was not measured.
Meader 2014[20,21]	SAQAT (Semi-Automated	Study limitations domain	**Study limitations domain – No serious limitations**: No problem for any source of risk of bias.

Article ID	Tool	Items	Response options
	Quality Assessment Tool)	1. Were data reported consistently for the outcome of interest (i.e. no potential selective reporting)? Publication bias domain 1. Did the authors conduct a comprehensive search? 2. Did the authors search for grey literature? 3. Authors did not apply restrictions to study selection on the basis of language? 4. There was no industry influence on studies included in the review? 5. There was no evidence of funnel plot asymmetry? 6. There was no discrepancy in findings between published and unpublished trials?	**Study limitations domain – Serious limitations**: Selection bias results in serious limitations, or very serious limitations if combined with a problem from any alternative source; two problems from other sources (e.g. detection bias, attrition bias) result in serious limitations. **Study limitations domain – Very serious limitations**: Selection bias results in serious limitations, or very serious limitations if combined with a problem from any alternative source; three problems result in very serious limitations **Publication bias domain – Strongly suspected**: High probability of publication bias. Responses to each item are entered into a Bayesian network to ascertain the probabilities of each GRADE domain. Publication bias is determined by a combination of discrepancy between published and unpublished studies (yes/no), amount of statistical information (high/intermediate/low), industry influence (yes/no) and search integrity (high/low), with the former carrying greatest weight. That is, the probability of publication bias is always considered high when there is a discrepancy between published and unpublished studies (regardless of responses to other items). **Publication bias domain – Undetected**: Low probability of publication bias (as determined by the Bayesian network described above.

Article ID	Tool	Items	Response options
Reid 2015[22]	Selective reporting bias algorithm	1. Protocol available? 2. Trial registration? 3. Outcomes described? 4. Response from contact with study authors? 5. Outcomes match?	**High risk of bias**: Outcomes are described in the protocol or trial registry or by the review authors when contacted, and they do not match the outcomes reported. **Low risk of bias**: Outcomes are described in the protocol or trial registry or by the review authors when contacted, and they do match the outcomes reported. **Unclear risk of bias**: Outcomes are not described in the protocol or trial registry, or a protocol or trial registry are not available and no response is received from review authors when contacted.
Saini 2014[23]	ORBIT-II (Outcome Reporting Bias In Trials) classification system for harm outcomes	1. ORBIT-II classification system	**Low risk of bias**: Specific examples include: (P3) Explicit specific harm measured and compared across treatment groups, although insufficient reporting for meta-analysis or full tabulation; (T1) Clinical judgement says specific harm likely measured but no events, because specific harm not mentioned but all other specific harms fully reported; (T2) Clinical judgement says specific harm likely measured but no events, because there was no description of specific harms; (U) Specific harm outcome not explicitly mentioned, clinical judgment says unlikely measured (no harms mentioned or reported). **High risk of bias**: In the context of harm outcomes, we awarded classifications for "high risk" outcome reporting bias when the specific harm had been measured but the data were presented or suppressed in a way that would mask the harm profile of particular interventions (including providing detail on the seriousness of the harms)—that is, P1, P2, R, and S classifications. Specific examples include: (P1) States outcome analysed but reported only that P>0.05; (P2) States outcome analysed

Article ID	Tool	Items	Response options
			but reported only that P<0.05; (R1) Clear that outcome was measured but no results reported; (R2) Result reported globally across all groups; (R3) Result reported from some groups only; (S1) Clinical judgment says specific harm outcome likely measured and likely compared across treatment groups, but only pooled adverse events reported (could include specific harm outcome); (S2) Clinical judgment says specific harm outcome likely measured and likely compared across treatment groups, but no harms mentioned or reported. **No risk of bias**: Specific examples include: (Q) Clear that explicit specific harm outcome was measured and clear outcome was not compared; (V) Report clearly specifies that data on specific harm of interest was not measured.
Salanti 2014[24 25]	Framework for evaluating the quality of evidence from a network meta-analysis	1. Study limitations (including selective outcome reporting) evaluated in a specific pairwise effect estimated in network meta-analysis: Determine which direct comparisons contribute to estimation of the NMA treatment effect and integrate risk of bias assessments from these into a single judgment. 2. Publication bias evaluated in a specific pairwise effect estimated in network meta-analysis: Non-statistical consideration of likelihood of non-publication of evidence that would inform the pairwise comparison. Plot pairwise estimates on contour-enhanced funnel plot. 3. Study limitations (including selective outcome reporting) evaluated in treatment ranking estimated in network meta-analysis: Integrate risk of bias assessments from each direct comparison to	**Study limitations domain – No serious limitations, do not downgrade**: Use standard GRADE considerations to inform judgment [7]. **Study limitations domain – Serious limitations, rate down one level (i.e., from high to moderate quality)**: Use standard GRADE considerations to inform judgment [7]. **Study limitations domain – Very serious limitations, rate down two levels (i.e., from high to low quality or moderate to very low)**: Use standard GRADE considerations to inform judgment [7]. **Publication bias domain (evaluated in a specific pairwise effect estimated in network meta-analysis) – Undetected**: Use standard GRADE to inform judgment [6].

Article ID	Tool	Items	Response options
		formulate a single overall confidence rating for treatment rankings. 4. Publication bias evaluated in treatment ranking estimated in network meta-analysis: Non-statistical consideration of likelihood of non-publication for each pairwise comparison. If appropriate, plot NMA estimates on a comparison adjusted funnel plot and assess asymmetry.	**Publication bias domain (evaluated in a specific pairwise effect estimated in network meta-analysis) – Strongly suspected**: "Even after a meticulous search for studies, publication bias can occur and usually it tends to lead to overestimation of an active treatment's effect compared with placebo or other reference treatment. Several approaches have been proposed to generate assumptions about the presence of publication bias, including funnel plots, regression methods and selection models, but each has limitations and their appropriateness is often debated. Making judgements about the presence of publication bias in a network meta-analysis is usually difficult. We suggest that for each observed pairwise comparison, judgements about the presence of publication bias are made using standard GRADE. We recommend that the primary considerations are non-statistical (by considering how likely it is that studies may have been performed but not published) and we advocate the use of contour-enhanced funnel plots, which may help in identifying publication bias as a likely explanation of funnel plot asymmetry. Then, judgements about the direct effects can be summarized to infer about the network estimates by taking into account the contributions of each direct piece of evidence" [24]. **Publication bias domain (evaluated in treatment ranking estimated in network meta-analysis) – Undetected**: Use standard GRADE to inform judgment [6]. **Publication bias domain (evaluated in treatment ranking estimated in network meta-analysis) – Strongly suspected**: "Judgments about the potential impact of

Article ID	Tool	Items	Response options
			publication bias in the ranking of the treatments require, as before, consideration of the comprehensiveness of the search for studies and the likelihood that studies may have been conducted and not published. A statistical approach to detecting bias is offered in certain situations by the comparison-adjusted funnel plot for a network of treatments. In such a plot, the vertical axis represents the inverted standard error of the effect sizes as in a standard funnel plot. However, the horizontal axis represents an adjusted effect size, presenting the difference between each observed effect size and the mean effect size for the specific comparison being made. The use of such a plot is informative only when the comparisons can confidently be ordered in a meaningful way; for example, if all comparisons are of active treatment versus placebo, or all are of a new versus an old drug. Examination of any asymmetry in the plot can help to infer about the possible presence of an association between study size and study effect. Asymmetry does not provide evidence of publication bias, however, since associations between effect size and study size can be due to study limitations or genuine heterogeneity of effects” [24].
Sterne 2016[26]	ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool	1. Is the reported effect estimate likely to be selected, on the basis of the results, from multiple outcome measurements within the outcome domain, multiple analyses of the intervention-outcome relationship, or different subgroups?	**Low risk of bias**: There is clear evidence (usually through examination of a pre-registered protocol or statistical analysis plan) that all reported results correspond to all intended outcomes, analyses and subcohorts. **Moderate risk of bias**: (i) The outcome measurements and analyses are consistent with an a priori plan; or are clearly defined and both internally and externally consistent; and (ii) there is no indication of selection of

Article ID	Tool	Items	Response options
			the reported analysis from among multiple analyses; and (iii) There is no indication of selection of the cohort or subgroups for analysis and reporting on the basis of the results. **Serious risk of bias**: (i) Outcomes are defined in different ways in the methods and results sections, or in different publications of the study; or (ii) There is a high risk of selective reporting from among multiple analyses; or (iii) The cohort or subgroup is selected from a larger study for analysis and appears to be reported on the basis of the results. **Critical risk of bias**: (i) There is evidence or strong suspicion of selective reporting of results; and (ii) The unreported results are likely to be substantially different from the reported results. **No information**: There is too little information to make a judgement (for example, if only an abstract is available for the study).
Viswanathan 2012[27]	RTI Item Bank for Assessment of Risk of Bias and Precision for Observational Studies of Interventions or Exposures	1. Are any important primary outcomes missing from the results? 2. Are any important harms or adverse events that may be a consequence of the intervention/exposure missing from the results?	**Yes (for item on primary outcome)**: No specific criteria stated. Only guidance is "Identify all primary outcomes, including timing of measurement, that one would expect to be reported in the study" **No (for item on primary outcome)**: No specific criteria stated. **Cannot determine (for item on primary outcome)**: No specific criteria stated. **Yes (for item on harm outcome)**: No specific criteria stated. Only guidance is "Identify all important harms,

Article ID	Tool	Items	Response options
			including timing of measurement, that one would expect be reported in the study. Drop if not relevant to body of literature." **Partially (for item on harm outcome)**: No specific criteria stated. **No (for item on harm outcome)**: No specific criteria stated. **Assessment of harms not applicable to this study (for item on harm outcome)**: No specific criteria stated.
Viswanathan 2013[28]	RTI Item Bank for Assessing Risk of Bias and Confounding for Observational Studies of Interventions or Exposures	1. Are any important primary outcomes missing from the results? 2. Are any important harms or adverse events that may be a consequence of the intervention/exposure missing from the results?	**Yes, important outcome(s) missing (for item on primary outcome)**: No specific criteria stated. Only guidance is "Identify all primary outcomes that one would expect to be reported in the study, including timing of measurement." **No important outcome (s) missing (for item on primary outcome)**: No specific criteria stated. **Cannot determine (for item on primary outcome)**: No specific criteria stated. **Yes, important outcomes missing (for item on harm outcome)**: No specific criteria stated. Only guidance is "Identify all important harms that one would expect be reported in the study, including timing of measurement. Drop if not relevant to body of literature." **No important outcomes missing (for item on harm outcome)**: No specific criteria stated. **Assessment of harms not applicable to this study (for item on harm outcome)**: No specific criteria stated.

References

1. Balshem H, Stevens A, Ansari M, et al. Finding grey literature evidence and assessing for outcome and analysis reporting biases when comparing medical interventions: AHRQ and the Effective Health Care Program. (Prepared by the Oregon Health and Science University and the University of Ottawa Evidence-based Practice Centers under Contract Nos. 290-2007-10057-I and 290-2007-10059-I.) AHRQ Publication No. 13(14)-EHC096-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

2. Berkman ND, Lohr KN, Ansari M, et al. Chapter 15 Appendix A: A Tool for Evaluating the Risk of Reporting Bias (in Chapter 15: Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update). Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I). AHRQ Publication No. 13(14)-EHC130-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2013. www.effectivehealthcare.ahrq.gov/reports/final.cfm

3. Downes MJ, Brennan ML, Williams HC, et al. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ open* 2016;6:e011458.

4. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.

5. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.

6. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64(12):1277-82.

7. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol* 2011;64(4):407-15.

8. Schünemann H, Brożek J, Guyatt G, et al. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. [Updated October 2013]. Available from http://gdt.guidelinedevelopment.org/app/handbook/handbook.html.

9. Santesso N, Carrasco-Labra A, Langendam M, et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. *J Clin Epidemiol* 2016

10. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.

11. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons 2008:187-241.

12. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org/.

13. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.

14. Higgins JPT, Savović J, Page MJ, et al. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0), Version 20 October 2016. Available from http://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/rob2-0/ [accessed 19 September 2017].

15. Higgins JPT, Sterne JAC, Savović J, et al. A revised tool for assessing risk of bias in randomized trials. *Cochrane Methods Cochrane Database of Systematic Reviews* 2016;10(Suppl 1):29-31.

16. Hooijmans CR, Rovers MM, de Vries RBM, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.

17. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013;66(4):408-14.

18. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.

19. Dwan K, Gamble C, Kolamunnage-Dona R, et al. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials* 2010;11:52.

20. Meader N, King K, Llewellyn A, et al. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic reviews* 2014;3(1):82.

21. Stewart GB, Higgins JP, Schunemann H, et al. The use of Bayesian networks to assess the quality of evidence from research synthesis: 1. *PLoS One* 2015;10(3):e0114497.

22. Reid EK, Tejani AM, Huan LN, et al. Managing the incidence of selective reporting bias: a survey of Cochrane review groups. *Systematic reviews* 2015;4:85.

23. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ* 2014;349:g6501.

24. Salanti G, Giovane CD, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014;9(7):e99682.

25. Higgins JP, Del Giovane C, Chaimani A, et al. Evaluating the quality of evidence from a network meta-analysis. *Value Health* 2014;17(7):A324.

26. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.

27. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol* 2012;65(2):163-78.

28. Viswanathan M, Berkman ND, Dryden DM, et al. AHRQ Methods for Effective Health Care. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or

Exposures: Further Development of the RTI Item Bank. Rockville (MD): Agency for Healthcare Research and Quality (US) 2013.

Table S5. General characteristics of studies evaluating the measurement properties of tools for assessing risk of reporting biases

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Armijo-Olivo 2012[1]	Cochrane risk of bias tool for randomized trials (2008 version)	None	20 trials included in a SR exploring knowledge transfer interventions for cancer pain management.	Cancer pain	None	20	NA	Range 1987-2007	2
Armijo-Olivo 2014[2]	Cochrane risk of bias tool for randomized trials (2011 version)	Inter-rater reliability	Trials of physical therapy interventions included in meta-analyses of a continuous outcome.	Physical therapy for musculoskeletal, cardiorespiratory, neurological or gynaecological conditions	None	109	NA	Not reported	2
Bilandzic 2016[3]	ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool	Inter-rater reliability	Studies included in two SRs of NRSI of the relationship between the use of TZDs and COX-2 inhibitors and major cardiovascular events.	Cardiovascular disease	None	37	NA	Range 2000-2010	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Downs 1998[4]	Downs-Black tool	None	10 randomised controlled trials and 10 non-randomised trials/prospective cohort studies randomly selected from studies identified during a SR of surgery for stress incontinence	Stress incontinence	None	20	NA	Not reported	2
Hartling 2009[5]	Cochrane risk of bias tool for randomized trials (2008 version)	Inter-rater reliability	A convenience sample of 163 randomized trial in child health, which were presented at the annual scientific meetings of the Society for Pediatric Research between 1992 and 1995.	Child health	None	163	NA	Not reported	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Hartling 2011[6]	Cochrane risk of bias tool for randomized trials (2008 version)	Inter-rater reliability	Trials included in a systematic review of long-acting beta agonists (LABA) combined with inhaled corticosteroids (ICS) for adults with persistent asthma.	Asthma	None	107	NA	Median 2004, IQR 2001-2006	2
Hartling 2012[7 8]	Cochrane risk of bias tool for randomized trials (2011 version)	Inter-rater reliability	A sample of 154 trial was randomly selected from among 616 trials published in December 2006 that were previously examined for quality of reporting.	Varied	None	154	NA	All 2006	2
Hayden 2013[9]	QUIPS (Quality In Prognosis Studies) tool	Inter-rater reliability	Studies included in a systematic review of troponin-based risk stratification of patients with	Pulmonary embolism	None	31	NA	Not reported	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
			acute non-massive pulmonary embolism.						
Hoojimans 2014[10]	SYRCLE's RoB tool (SYstematic Review Centre for Laboratory animal Experimentation)	Inter-rater reliability	1 systematic review including 32 papers (no other details provided).	Animal studies (not specified)	None	32	NA	Not reported	2
Jordan 2017[11]	Cochrane risk of bias tool for randomized trials (2011 version)	Inter-rater reliability	Any study that had been included more than once in SRs present on the Cochrane Database of Systematic Reviews in the area of subfertility.	Subfertility	None	28	NA	Not reported	2
Kim 2013[12]	RoBANS (Risk of Bias Assessment Tool for Nonrandomized Studies)	Inter-rater reliability	39 NRSs from four systematic reviews (one by the National Evidence-based Healthcare Collaborating Agency and three Cochrane reviews).	Depression, myocardial infarction, post-partum hemorrhage, chronic non-cancer pain	None	39	NA	Not reported	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Kumar 2016[13]	GRADE	None	10 key questions that were systematically reviewed for a clinical practice guideline for the use of prophylactic vs. therapeutic platelet transfusion in patients with thrombocytopenia.	Thrombocytopenia	10	None	April 2015	NA	18
Llewellyn 2015[14]	SAQAT (Semi-Automated Quality Assessment Tool)	Inter-rater reliability	29 meta-analyses from a purposive sample of SRs of RCTs from the Database of Systematic Reviews of Effects (DARE), and a purposive sample of 15 recent Cochrane reviews in mental health.	Varied	44	None	2006-2013	NA	2
Mustafa 2013[15]	GRADE	None	4 well-conducted and well-reported Cochrane reviews,	Alcohol dependence, asthma,	16	None	2004-2012	NA	4

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
			based on assessment using the AMSTAR tool.	cardiopulmonary bypass					
Norris 2012[16]	ORBIT-I (Outcome Reporting Bias In Trials) classification system for benefit outcomes	Inter-rater reliability; Time to complete assessments	Studies included in three AHRQ-funded comparative effectiveness reviews of randomised trials with drug-drug or drug-placebo comparisons, examining benefit outcomes.	Varied	None	40	NA	2005-2010	2
O'Connor 2015[17]	Downs-Black tool	None	20 studies included in an updated SR which examined the effects of an exercise intervention for chronic musculoskeletal pain.	Chronic musculoskeletal pain	None	20	NA	1997-2008	2

Study ID	Tool assessed	Properties evaluated for reporting bias item	Sampling frame	Areas of health care	No. syntheses assessed	No. studies assessed	Publication years of syntheses	Publication years of studies	No. assessors
Vale 2013[18]	Cochrane risk of bias tool for randomized trials (2011 version)	Agreement between assessments performed using published article only versus published article and data collected during the individual participant data process.	13 completed individual participant data meta-analyses of treatments for cancer. Trials had to be published either in full or as an abstract, and a copy of the trial protocol or forms detailing trial design completed by trialists (or both) had to be available.	Cancer pain	None	95	NA	Not reported	2

NA = Not applicable; SR = systematic review

References

1. Armijo-Olivo S, Stiles CR, Hagen NA, et al. Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. *J Eval Clin Pract* 2012;18(1):12-8.

2. Armijo-Olivo S, Ospina M, da Costa BR, et al. Poor reliability between Cochrane reviewers and blinded external reviewers when applying the Cochrane risk of bias tool in physical therapy trials. *PLoS One* 2014;9(5):e96920.

3. Bilandzic A, Fitzpatrick T, Rosella L, et al. Risk of Bias in Systematic Reviews of Non-Randomized Studies of Adverse Cardiovascular Effects of Thiazolidinediones and Cyclooxygenase-2 Inhibitors: Application of a New Cochrane Risk of Bias Tool. *PLoS Med* 2016;13(4):e1001987.

4. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.

5. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ* 2009;339:b4012.

6. Hartling L, Bond K, Vandermeer B, et al. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. *PLoS One* 2011;6(2):e17242.

7. Hartling L, Hamm M, Milne A, et al. AHRQ Methods for Effective Health Care. Validity and Inter-Rater Reliability Testing of Quality Assessment Instruments. Rockville (MD): Agency for Healthcare Research and Quality (US) 2012.

8. Hartling L, Hamm MP, Milne A, et al. Testing the risk of bias tool showed low reliability between individual reviewers and across consensus assessments of reviewer pairs. *J Clin Epidemiol* 2013;66(9):973-81.

9. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6.

10. Hooijmans CR, Rovers MM, de Vries RBM, et al. SYRCLE's risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.

11. Jordan VM, Lensen SF, Farquhar CM. There were large discrepancies in risk of bias tool judgments when a randomized controlled trial appeared in more than one systematic review. *J Clin Epidemiol* 2017;81:72-76.

12. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J Clin Epidemiol* 2013;66(4):408-14.

13. Kumar A, Miladinovic B, Guyatt GH, et al. GRADE guidelines system is reproducible when instructions are clearly operationalized even among the guidelines panel members with limited experience with GRADE. *J Clin Epidemiol* 2016;75:115-8.

14. Llewellyn A, Whittington C, Stewart G, et al. The Use of Bayesian Networks to Assess the Quality of Evidence from Research Synthesis: 2. Inter-Rater Reliability and Comparison with Standard GRADE Assessment. *PLoS One* 2015;10(12):e0123511.

15. Mustafa RA, Santesso N, Brozek J, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol* 2013;66(7):736-42; quiz 42.e1-5.

16. Norris SL, Holmer HK, Ogden LA, et al. AHRQ Methods for Effective Health Care. Selective Outcome Reporting as a Source of Bias in Reviews of Comparative Effectiveness. Rockville (MD): Agency for Healthcare Research and Quality (US) 2012.

17. O'Connor SR, Tully MA, Ryan B, et al. Failure of a numerical quality assessment scale to identify potential risk of bias in a systematic review: a comparison study. *BMC Res Notes* 2015;8:224.

18. Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. *BMJ* 2013;346:f1798.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Table S1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7-8
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	NA
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	NA

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	NA
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8, Fig 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	12
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	NA
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Table S3 and S4
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	NA
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	13-22
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	23
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	23-24
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	26
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	26-27

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.