



RESEARCH ARTICLE

REVISED Toolkit of methodological resources to conduct systematic reviews [version 2; peer review: 2 approved]

Marta Roqué ^{1,2}, Laura Martínez-García ^{1,2}, Ivan Solà^{1,2}, Pablo Alonso-Coello^{1,2}, Xavier Bonfill¹⁻³, Javier Zamora^{2,4}

¹Iberoamerican Cochrane Centre - Sant Pau Biomedical Research Institute (IIB-Sant Pau), Barcelona, Spain

²CIBER of Epidemiology and Public Health (CIBERESP), Madrid, Spain

³Autonomous University of Barcelona, Bellaterra, Spain

⁴Clinical Biostatistics Unit, Ramón y Cajal Health Research Institute, Madrid, Spain

V2 First published: 04 Feb 2020, 9:82
<https://doi.org/10.12688/f1000research.22032.1>

Latest published: 11 Aug 2020, 9:82
<https://doi.org/10.12688/f1000research.22032.2>

Abstract

Background: Systematic reviews (SR) can be classified by type depending on the research question they are based on. This work identifies and describes the most relevant methodological resources to conduct high-quality reviews that answer health care questions regarding prevalence, prognosis, diagnostic accuracy and effects of interventions.

Methods: Methodological resources have been identified from literature searches and consulting guidelines from institutions that develop SRs. The selected resources are organized by type of SR, and stage of development of the review (formulation of the research question, development of the protocol, literature search, risk of bias assessment, synthesis of findings, assessment of the quality of evidence, and report of SR results and conclusions).

Results: Although the different types of SRs are developed following the same steps, each SR type requires specific methods, differing in characteristics and complexity. The extent of methodological development varies by type of SR, with more solid guidelines available for diagnostic accuracy and effects of interventions SRs.

This methodological toolkit describes the most up-to-date risk of bias instruments: Quality in Prognostic Studies (QUIPS) tool and Prediction model study Risk Of Bias Assessment Tool (PROBAST) for prognostic SRs, Quality assessment of diagnostic accuracy studies tool (QUADAS-2) for diagnostic accuracy SRs, Cochrane risk of bias tool (ROB-2) and Risk of bias in non-randomised studies of interventions studies tool (ROBINS-I) for effects of interventions SRs, as well as the latest developments on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system.

Conclusions: This structured compilation of the best methodological resources for each type of SR may prove to be a very useful tool for those researchers that wish to develop SRs or conduct methodological research works on SRs

Open Peer Review

Reviewer Status

	Invited Reviewers	
	1	2
version 2		
(revision)		
11 Aug 2020	report	report
	↑	↑
version 1		
04 Feb 2020	?	?
	report	report

1. **Miranda Cumpston** , Monash University, Melbourne, Australia
University of Newcastle, Newcastle, Australia
2. **Edward Purcell** , City, University of London, London, UK

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

Systematic reviews, prevalence, prognostic, diagnostic accuracy, efficacy of interventions

Corresponding author: Marta Roqué (mroque@santpau.cat)

Author roles: **Roqué M:** Conceptualization, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Martínez-García L:** Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Solà I:** Writing – Original Draft Preparation; **Alonso-Coello P:** Writing – Original Draft Preparation; **Bonfill X:** Writing – Original Draft Preparation; **Zamora J:** Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: The authors are members of CIBERESP (Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública - Biomedical Research Center Network of Epidemiology and Public Health), and hold active roles within Cochrane and the GRADE Working Group.

Grant information: Roqué M is supported by the Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP) as part of a Training Programme call for “Internal mobility: Internships in CIBERESP groups”, within the framework of the subprogramme 7.4 “Methodology, clinical records and scientific dissemination.” Martínez-García L has a Miguel Servet contract from the Institute of Health Carlos III [CP18/00007].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Roqué M *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Roqué M, Martínez-García L, Solà I *et al.* **Toolkit of methodological resources to conduct systematic reviews [version 2; peer review: 2 approved]** F1000Research 2020, 9:82 <https://doi.org/10.12688/f1000research.22032.2>

First published: 04 Feb 2020, 9:82 <https://doi.org/10.12688/f1000research.22032.1>

REVISED Amendments from Version 1

This new version incorporates a number of clarifications and modifications suggested by the peer reviewers, importantly a table organizing the resources identified in this project, classified into guidelines to conduct systematic reviews, reporting guidelines, primary methods papers and references to examples used in the manuscript.

Any further responses from the reviewers can be found at the end of the article

Introduction

Systematic reviews (SR) are studies that use a systematic and explicit method to identify, analyse and synthesize empirical evidence, and to answer a specific research question¹. Therefore, SRs are key tools to make informed health choices^{2,3}.

All SRs are based on a specific research question. Classic epidemiological research questions relate to the prevalence of a medical condition, the associated prognosis of the medical condition (including incidence or global prognosis, prognostic factors associated to the condition's incidence or outcome, and risk profiles defined by prognostic models⁴), diagnostic accuracy of

tests that allow us to diagnose the medical condition, and effects of interventions to treat the medical condition. SRs can be classified by the type of research question they answer, as shown in [Table 1](#).

The stages to develop an SR are common to all the types of SRs: 1) Formulating the research question, 2) development of the protocol that explicitly describes the methods to carry out each step of the SR, 3) literature search, 4) risk of bias assessment, 5) synthesis of findings, 6) assessment of the quality of evidence, and 7) report of SR results and conclusions¹. Although the different types of SRs share the same structure and follow a similar development process, their methods can be different and more or less complex depending on the type of SR.

Nowadays there are numerous methodological resources to conduct reviews, especially for intervention SRs and diagnostic SRs. However, the scattering of these resources and the lack of widely established manuals or recommendations are, in many situations, an obstacle to access them, especially for prevalence SRs and prognostic SRs. Therefore, the objective of this review is to identify and describe the methodological resources available to develop prevalence SRs, prognostic SRs, diagnostic accuracy SRs and effects of interventions SRs.

Table 1. Research question by type of systematic review.

Type of systematic review	Acronym for the research question	Example of research question
Prevalence review	CoCoPop-S (<i>condition, context, population and study design</i>)	What is the prevalence of frailty and prefrailty (condition) in community-dwelling older adults (population) living in low- and middle-income countries (context)? ⁵ What is the worldwide (population) prevalence of insufficient physical activity (condition)? ⁶
Prognostic review - global prognosis	CoCoPop-S (<i>condition, context, population and study design</i>)	What is the incidence of dementia (condition) in individuals of at least 60 years of age (population) living in high-income countries (context)? ⁷
Prognostic review- prognostic factors	PICOT-S (<i>population, intervention or factor, comparison, outcome, time and study design</i>) PFO-S (<i>population, factor or model, outcome and study design</i>)	Is protease activity (prognostic factor) an independent prognostic factor for wound healing (outcome) at 24 weeks (timeframe) in people with venous leg ulcers (population)? ⁸
Prognostic review- prognostic models	PICOT-S (<i>population, intervention or factor, comparison, outcome, time and study design</i>)	What is best prognostic model to predict overall or progression- free survival (outcome) in patients with chronic lymphocytic leukaemia (condition)? ⁹
Diagnostic accuracy review	PIRD-S (<i>population, index test, reference test, diagnosis of interest and study design</i>)	Do self-reported frailty to predict survival in adults with bacterial meningitis screening instruments (index test) accurately identify older people (population) at risk of frailty and prefrailty (condition of interest)? ¹⁰ Is PET 18F florbetapen (index test) useful in early diagnosing dementia (condition) in patients with mild cognitive impairment (population)? ¹¹
Effects of intervention review	PICO-S (<i>population, intervention, comparison, outcome of interest and study design</i>)	What is the effect of ribavirin (intervention) in patients with Crimean Congo haemorrhagic fever to prevent death (outcome)? ¹² Does comprehensive geriatric assessment (intervention) in older adults (population) reduce mortality (outcome)? ¹³

Methods

Information sources and search strategy

We consulted the guidelines from the main organizations that establish methods to conduct SRs (Cochrane, Joanna Briggs Institute, European Network for Health Technology Assessment (EUNETHTA), Enhancing the Quality and Transparency of Health Research (EQUATOR) network, Grading of Recommendations Assessment, Development and Evaluation (GRADE)) in order to identify their proposed resources.

Additionally, we performed a literature search in MEDLINE (accessed through PubMed) in November 2019 using the following search syntax: (("Review Literature as Topic"[Mesh] OR systematic review*[tiab]) AND (handbook*[ti] OR methodolog*[ti] OR manual[ti] OR guide[ti]).

We also performed ad hoc scientific literature searches to find other resources for each type of SR in relation to the research question structure, the literature search strategy, the risk of bias assessment and the statistical analysis.

Eligibility criteria

We included the resources available to design prevalence SRs, prognostic SRs, diagnostic SRs and intervention SRs.

We excluded the methodological resources to develop other types of SRs (methodological, economic evaluation and qualitative research SRs, or overviews).

Data selection and extraction

The authors are members of CIBERESP (Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública - Biomedical Research Center Network of Epidemiology and Public Health), hold active roles within Cochrane and the GRADE Working Group, and are experts in different fields of knowledge (statistics, development of Cochrane reviews, research methodology, information retrieval, development of clinical guidelines). They evaluated the search results, selected the most relevant and accurate resources, and summarized the most relevant information by development stage and type of SR.

The resources were selected based on the authors expert judgement, prioritising those resources which were endorsed or part of a guideline from the organisations cited above, and those which were more recent. The resources were organised in 7 sections, following the development stages of an SR: 1) Formulating the research question, 2) development of the protocol and review registration, 3) search strategy, 4) risk of bias assessment, 5) statistical synthesis of findings, 6) quality of evidence assessment, and 7) results report and presentation. The resources are presented by type of SR in each section, and an example of their use is included⁵⁻¹³.

For each pre-defined section, the authors selected and summarized the methods that were considered to be more rigorous and widely accepted, prioritizing major methods applicable to all reviews over more controversial methods, or methods which required highly specialized knowledge. The text organises the

results pedagogically with the aim to highlight key differences between review types, present the key characteristics of each method, and be a comprehensive tool that contains the most relevant advice based on the authors judgement.

Results

We identified guidance handbooks, primary studies and reporting guidelines as a result of the bibliographic searches. The resources selected are presented in Table 2.

We have identified methodological guidelines dedicated to the development of prevalence SRs¹⁴, global prognosis¹⁵, and prognostic factor SRs¹⁶⁻¹⁸.

During the performed search, we identified methodological manuals to develop prognostic model SRs in the series of publications from the PROGRESS project¹⁹, and in the resource compilation from Cochrane's Prognosis Methods Group.

For diagnostic accuracy SRs and effects of interventions SRs, we have identified the methodological manuals developed by Cochrane Collaboration are available^{1,20}. The recommendations drawn from the guidance handbooks identified are complemented whenever necessary with specific primary method studies identified in our search.

Formulating the research question

The type of SR is determined by the **research question**, which must be formulated in a structured manner as shown in Table 1. Careful development of the research question is vital, since the SR inclusion criteria will stem from it.

Prevalence review. Prevalence SRs aim to answer the question "How common is a health problem in a specific population?" Prevalence SRs focus on existing cases at a given time, measure the global burden of a health problem, and describe the characteristics of the affected population, the geographical distribution of that problem and its variation among subgroups. The structure of the research question must include the elements of condition, context, population and study design (CoCoPop-S)²¹, as shown in Table 1. The most adequate study designs to estimate the prevalence would be population registers or cross-sectional studies that include population-representative samples. For instance, Guthold *et al.* (2018) considers studies based on population surveys as a reliable source of information to obtain global prevalence estimators of insufficient physical activity⁶.

Prognostic review. SRs of prognosis are mainly based on three types of research questions: 1) "What is the risk of an specific population to have a health problem?", descriptive question (review of global prognosis) that focuses in new cases occurring within a period of time (incidence), 2) "what factors are associated with or determine a specific outcome?", an explanatory question (review of prognostic factors), and 3) "are there risk profiles that have higher probability of presenting specific outcomes?", a result prediction question (review of prognostic models or risk prediction). We have excluded from the aim of this project a 4th type of prognostic question, known as

Table 2. Organisation of resources by type.

Best practice manuals and chapters of manuals	Primary methods	Reporting guidelines
JB1 manual	Hemingway 2013 ⁴	Moher 2009 ⁷²
Aromataris 2017 ¹⁴	Munn 2018 ²¹	Moher 2015 ⁷³
(Munn 2017 ¹⁵	Iorio 2015 ²²	Beller 2013 ⁷⁴
Moola 2017 ¹⁶	Bossuyt 2006 ²³	Zorzela 2016 ⁷⁵
Campbell 2017 ³¹)	Lijmer 1999 ²⁵	McInnes 2018 ⁷⁷
Cochrane DTA manual	Strauss 2010 ²⁶	Moher 2007 ⁷⁸
Deeks 2010 ²⁰	Ge 2018 ²⁷	Page 2016 ⁷⁹
(Bossuyt 2008 ²⁴	Page 2018 ²⁸	Salameh 2019 ⁸⁰
deVet 2008 ³²	Atkinson 2015 ³⁰	
Macaskill 2010 ⁶⁰)	Lefebvre 2013 ³³	
Cochrane intervention manual	Glanville 2006 ³⁴	
Higgins 2019 ¹	Wilczynski 2004 ³⁵	
(Lefebvre 2019 ²⁹	Beynon. 2013 ³⁶	
Higgins 2019 ⁵⁵	Sampson 2011 ³⁷	
Deeks 2019 ⁵⁸	Bramer 2017 ³⁸	
Chaimani 2019 ⁶³	Hartling 2016 ³⁹	
Schünemann 2019 ⁶⁸)	Glanville 2014 ⁴⁰	
GRADE Working group manual	Isojarvi 2018 ⁴¹	
Schünemann 2013 ⁶⁴	Horsley 2011 ⁴²	
(Schünemann 2020 ⁶⁵	Gentles 2016 ⁴³	
Schünemann 2020 ⁶⁶	Hartling 2017 ⁴⁴	
Santeso 2019 ⁶⁷)	Booth 2016 ⁴⁵	
PROGRESS project	Rethlefsen 2014 ⁴⁶	
Riley 2019 ¹⁷	Rethlefsen 2015 ⁴⁷	
Debray 2017 ¹⁹	Spencer 2018 ⁴⁸	
Dekkers 2019 ¹⁸ manual	Hoy 2012 ⁴⁹	
	Hayden 2013 ⁵⁰	
	Morgan 2018 ⁵¹	
	Morgan 2019 ⁵²	
	Wolff 2019 ⁵³	
	Whiting 2011 ⁵⁴	
	Sterne 2016 ⁵⁶	
	Lau 1997 ⁵⁷	
	Popay 2006 ⁵⁹	
	Rutter 2001 ⁶¹	
	Rücker 2008 ⁶²	
	Murad 2017 ⁶⁹	
	Harder 2017 ⁷⁰	
	Huguet 2013 ⁷¹	
	Campbell 2020 ⁷⁶	

stratified medicine, and that alludes to the use of prognostic information to individualise therapeutic choices in a group of people with similar characteristics⁴.

Structured questions about global prognosis must specify population, outcome, condition to be predicted, context and time frame to determine the incidence (CoCoPop-S). The study

designs that provide more reliable incidence estimates are prospective cohort studies with representative samples^{15,22}. Structured questions regarding either prognostic factors or models must include population; exposure in terms of the prognostic factor or model of interest, including how it is measured, the intensity and the exposure time; outcome, condition to be predicted; follow-up time; and context (PICOT-S or PFO-S)^{19,21}. The

best study designs to evaluate prognostic factors or models are also prospective cohort studies. For instance, Westby *et al.* (2018) published a prognostic factor SR that gives priority to the inclusion of cohort studies and, if none is found, it resorts to including case-control studies, which also explore the association of prognostic factors with the outcome of interest, although with less reliability⁸.

Diagnostic accuracy review. Diagnostic SRs aim to answer the question “How good is a test to identify or dismiss the presence of a condition or health problem in a particular population, in comparison with a reference test?” The research question can be posed with the elements of population, index test, reference test, diagnosis of interest and study design (PIRD-S)²¹. The SR approach will depend on the role of the index test in the clinical diagnostic pathway: if it replaces another test, if it will be used in addition to another test to refine the diagnosis, or if it is a triage test previous to other tests^{23,24}.

Diagnostic SRs preferentially include cross-sectional studies, where the participants are evaluated using the index test and/or the reference test to determine if they have the condition of interest. Case-control designs are subject to risk of bias and their inclusion in diagnostic SRs is not recommended²⁵. For instance, Ambagtsheer *et al.* (2017) include in their SR cross-sectional studies where one or more self-reported frailty screening scales have been compared with one of three reference standards: frailty phenotype, frailty index or comprehensive geriatric assessment¹⁰.

Effects of interventions review Interventions SRs aim to answer the question “What effect does a specific intervention have on the relevant outcomes in people with a particular health problem, in comparison with a reference intervention?” The research question is posed with the elements of population, intervention, comparator, outcomes of interest and study design (PICO-S)¹.

The randomised clinical trial (RCT) is the most appropriate study design to evaluate the effects of an intervention, as it is the design with less risk of bias and that best helps to establish causality. In cases where it is not possible to conduct randomised trials for ethical or organizational reasons, non-randomised trials, before-after studies, time series, cohort studies or case-control studies can be considered for their inclusion in the SR¹. For instance, the SR by Johnson *et al.* (2018) regarding ribavirin for treating Crimean Congo haemorrhagic fever included both RCTs and non-randomised trials to use the available data, given the previous lack of preparedness for experimental research therapeutics in outbreak situations, but concludes that estimates of effect based on the existing literature are highly uncertain due to confounding in non-randomised studies¹².

Development of the protocol and review registration

Writing the SR protocol is a fundamental step that must be done before designing an SR. Herein, the stages and methods to be applied during the development of the SR can be pre-specified. The identified guidelines can be used to identify the methods that need to be stated in the protocol, and some have specific chapters on protocol development^{1,14,20}.

Similarly to the requirement of clinical trial registration, the SR should also be registered in order to avoid redundancies and, more importantly, to avoid reporting bias, therefore guaranteeing transparency and rigor during the development of the SR²⁶. Prospective registration of an SR protocol is recommended by the PRISMA guidelines and is associated with higher SR methodological quality²⁷. The largest and most well-known SR register is **PROSPERO**, produced by the Centre for Reviews and Dissemination in York. With PROSPERO, it is possible to prospectively register any type of review, provided that its aim is a health-related outcome. It contains more than 30,000 entries²⁸. All Cochrane SR protocols are published in Cochrane Library and automatically registered in PROSPERO.

Search strategy

Designing a comprehensive research study for an SR is vital in order to reduce bias when identifying studies, and it is important to describe it in the relevant section within the protocol in a transparent and thorough manner to facilitate its evaluation by third parties and its reproducibility.

Methodological reference standards to design comprehensive searches have been published^{29,30}. In addition, methodological manuals to develop SRs provide guidelines for diagnostic and effects of interventions SRs^{31–33}.

The design of the search strategies does not differ by type of SR, but rather their differences are due to the elements of the research question and the design of studies to be identified. In general terms, electronic searches are designed to identify bibliographic references that use a language similar to the elements of the review’s clinical question. To this effect, the strategies are built based on the elements of the structured clinical question. Search algorithms use a combination of natural language and the appropriate controlled vocabulary for each bibliographic database. Validated filters can be applied to these strategies to determine specific study designs that can be useful to identify, among others, clinical trials^{32–34}, or prognostic studies³⁵. However, the use of filters is controversial in diagnostic accuracy studies^{32,36}.

Search performance will vary depending on the type of studies that are included in the SR. Thus, in intervention SRs, the search results for RCTs are more precise (they have a higher proportion of relevant references among all the references that the search has identified), due to better indexation of this type of studies in bibliographic databases. On the contrary, in SRs that include observational studies, like prognostic SRs, identifying studies is more complex given the variability of designs to be included and its poorer indexation in databases, which results in less specific literature searches that lead to a longer and more complex study selection process¹⁷.

Searches must be designed to optimise their sensitivity (the ability to retrieve as many relevant study references as possible), which is a feature that tends to be a detriment to precision, which in SRs ranges on an average of 3%³⁷. To obtain an efficient search with adequate sensitivity, performing searches in **MEDLINE** and **EMBASE** may be sufficient, particularly in intervention reviews, as they are the two most frequently used bibliographic

databases³⁸, and they are enough to identify most relevant studies for a specific SR³⁹. These searches can be complemented with additional searches in other databases such as PEDro, which provide specific information for certain topics.

Searching in bibliographic databases can be completed with additional strategies, such as checking public trial registers^{40,41}, searching in the reference list of relevant studies⁴², or cross-searching citations⁴³. Searching grey literature, understood as any document that is not published in biomedical or scientific journals, has a limited impact in effects of interventions SRs⁴⁴, but offers good results in other types of SRs, such as qualitative evaluation SRs⁴⁵.

If we take into consideration the methodological and technical challenges that the design and implementation of search strategies pose, involving a medical librarian can be desirable to improve the search quality^{46–48}.

Risk of bias assessment

Assessing the risk of bias of the included studies is a key element in any SR. It helps evaluate and interpret the included studies results, and it is a determinant of the evidence quality of the SR results. The current tools to assess risk of bias are organised by domains, which roughly correspond to the classic epidemiological biases related to each type of research question. The identified tools to assess risk of bias are presented in Table 3, organised by type of SR and by domain of epidemiological bias assessed.

Each of the domains of these tools includes a number of index questions related to specific aspects of study design or development that can lead to a bias in that domain. The tools can be adapted *a priori* to each review, modifying or deleting questions, or adding new questions specific to the considered research question. The process to assess risk of bias is similar in all the current scales. Firstly, they identify the risk of bias in each domain based on the answers to the questions, and secondly, they integrate these risks in a risk of bias assessment for each health problem, prognostic factor, diagnosed condition or outcome of interest assessed, depending on the type of SR.

Prevalence review. The tool to assess risk of bias by Hoy *et al.* (2012) is available for prevalence SRs. It assesses internal and external validity aspects in the prevalence study⁴⁹. The tool comprises 10 questions where a judgement of high or low risk of bias is made. Based on the answers, the researcher makes a subjective assessment of the study's overall risk of bias as low, moderate or high⁴⁹.

Prognostic review. There is no scale available to assess the risk of bias in **global prognostic studies**, although a series of criteria has been proposed to assess risk of bias. These are classified in 1) definition and representativeness of the population, 2) completeness of follow-up, and 3) objective and unbiased measurement of outcome of interest²². However, some authors like Roerh *et al.* (2018) use a version of the scale to assess risk of bias designed by Hoy *et al.* (2012), adapted to the assessment

of incidence studies considering the duration of the incidence period⁷.

For the **prognostic factor** studies, the tools QUIPS and “RoB instrument for NRS of exposures” were identified^{50–52}. The QUIPS tool helps assess the risk of bias using 31 questions divided in 6 domains. For each domain, a judgement of high, low or unclear risk of bias is made. Before using the tool, one must carefully consider the potential confounders that can lead to bias. Clinical experts in the specific topic of the SR should participate. The tool “RoB instrument for NRS of exposures” evaluates the risk of bias using 32 questions divided in 7 domains, including a key domain regarding confounders and a domain regarding departures from intended exposures. For each domain, a judgement of critical, serious, moderate or low risk of bias is made. An example of the use of the QUIPS scale can be seen in the review by Westby *et al.* (2018). The authors defined *a priori* two key confounders (age and infection), which the experts and the literature described as prognostic factors for their condition of interest (venous leg ulcers), and which were simultaneously associated with the prognostic factor of interest in the SR (protease activity biomarker). These two confounders were included in the QUIPS scale in the section of control by confounders⁸.

We identified the Prediction model Risk Of Bias ASsessment Tool (PROBAST) for the **prognostic** model SRs⁵³. This tool assesses the risk of bias using 20 questions divided in 4 domains (participants, predictors, outcome and analysis). For each domain, a judgement of high, low or unclear risk of bias is made. The questions vary according to the aim of the study (development, validation, or development and validation of the prognostic model).

Diagnostic accuracy review The tool QUADAS-2, which evaluates 11 questions divided in 4 domains, is available to assess the risk of bias in diagnostic accuracy studies⁵⁴. For each domain, a judgement of high, low or unclear risk of bias is made. In addition, the external validity or study applicability in relation to the SR is assessed in each domain.

Diagnostic SRs mainly include observational studies, which are more subject to risk of bias, and therefore adapting the QUADAS-2 tool, modifying or adding specific questions to the SR topic, is virtually a requirement during the protocol stage. For instance, the SR by Martínez *et al.* (2017) studied the diagnostic accuracy of an imaging test (amyloid PET) that requires complex visual interpretation. For this reason, a question was included in the QUADAS scale to assess whether the test interpretation was performed by trained readers¹¹.

Effects of interventions review. For intervention SRs, the Risk of Bias (RoB) 2.0 tool is available to assess the potential bias in randomised clinical trials, and the Risk Of Bias In Non-randomised Studies - of Interventions (RoBiNS-I) tool in non-randomised clinical trials^{55,56}. The RoB 2.0 tool includes 16 questions divided in 5 domains, including a specific domain for randomisation and a domain for deviations from intended

Table 3. Tools to assess risk of bias by type of systematic review.

	Scale (n items)	Selection bias (number of items)	Exposure and performance bias (number of items)	Outcome detection bias (number of items)	Attrition bias(number of items)	Confounder bias (number of items)	Selective outcome reporting bias (number of items)	Other biases (number of items)
Prevalence review	Hoy 2012 (10) ⁴⁹	- Representativeness of population sample (1) - Sample and recruitment (2)	(0)	- Data collection (2) - Case definition and timeframe for prevalence (2) - Reliability of measuring instrument (1)	- Impact of missing data (1)	(0)	(0)	- Appropriate computation of prevalence estimator (1)
Prognostic review- prognostic factors	QUIPS (31) ⁵⁰	- Study participation (3) - Sample and recruitment (3)	- Prognostic factors definition and measurement (6) - Confounders definition and measurement (4)	- Outcome definition and measurement (3)	- Description and impact of attrition (6)	- Statistical analysis of confounding factors (2)	- Selective reporting of results (1)	- Statistical analysis (3)
	RoB for NRS - exposures (32) ⁵²	- Selection of participants (5)	- Exposure definition and measurement (5) - Deviations from intended exposure (4)	- Outcome definition and measurement (5)	- Description and impact of attrition (5)	- Statistical analysis of confounding factors (6)	- Selective reporting of results (3)	(0)
Prognostic review- prognostic models	PROBAST (20) ⁵³	- Design of study and selection of participants (2)	- Prognostic factors definition and measurement (3)	- Outcome definition and measurement (6)	- Inclusion of participants in the analysis (2)	(0)	- Selective reporting of results (1)	- Statistical analysis (6)
Diagnostic accuracy review	QUADAS-2 (11) ⁵⁴	- Selection of participants (3)	- Index test interpretation (1) - Threshold specification for index test (1)	- Adequacy and interpretation of reference test(2) - Time interval between tests, and coverage of reference test (3)	- Inclusion of participants in the analysis (1)	(0)	(0)	(0)
Effects of intervention review	ROB-2 (16) ⁵⁵	- Selection of participants (randomisation, concealment, and basal imbalances) (3)	- Blinding of participants and personnel (2) - Deviations from intended intervention (2)	- Blinding of outcome detection (2)	- Impact of attrition (3)	(0)	- Selective reporting of results (2)	- Analysis of participants in the allocated intervention arm (2)
	ROBINS-I (35) ⁵⁶	- Selection of participants (6)	- Classification of intervention (3) - Deviations from intended intervention (6)	- Outcome measurement (4)	- Description and impact of attrition (5)	- Confounders (8)	- Selective reporting of results (3)	(0)

No risk of bias tool has been identified for global prognosis systematic reviews. The number of items in the risk of bias tools may vary depending on the effect of interest and the included study designs, as well as the addition or suppression of index questions by the researchers to tailor the tool to the SR.

interventions⁵⁵. The number of questions may vary, depending on the effect of interest and the design of the study assessed. For each domain, a judgement is made: high or low risk of bias, or some concerns. For instance, in their SR, Ellis *et al.* (2017) assessed the risk of bias in the evaluation of results separately for the objective outcomes (such as living at home or death) and for the subjective outcomes, showing a lower risk of bias in the evaluation of the objective outcomes¹³.

The RoBiNS-I tool assesses the biases that the non-randomised study has when compared with an ideal, pragmatic, unbiased randomised trial, which answers the clinical question of interest (even if this ideal trial may not be feasible or ethical)⁵⁶. RoBiNS-I has 34 questions divided in 7 domains, including a key domain regarding confounders and a domain for deviations from intended interventions. As in the case of prognostic SRs, there should be an *a priori* careful consideration of the potential confounders that must be included in the tool to assess individual studies. A judgement of critical, serious, moderate or low risk of bias is made for each domain. A low risk of bias implies that the non-randomised study is comparable to a well-performed randomised trial. For instance, Johnson *et al.* (2018) excluded from their analyses the non-randomised studies that showed a critical risk of bias according to RoBiNS-I, rejecting 18 out of the 22 included studies¹².

Statistical synthesis of findings

SRs may include a section with a quantitative statistical synthesis or meta-analysis, where a combined estimator of the

parameter of interest is obtained from the estimators of the individual studies. Table 4 shows a non-exhaustive compilation of the main characteristics of the meta-analysis methods and the main software commands for each type of SR.

A necessary previous step to any meta-analysis is the evaluation of the existing clinical and statistical heterogeneity in the set of studies, which will inform us 1) if it is reasonable to perform a quantitative synthesis of findings, 2) what meta-analysis model we should apply, and 3) if additional investigation of the causes of heterogeneity is required, for example, subgroup and sensitivity analyses, or meta-regressions^{57,58}. In those cases when a quantitative synthesis is precluded, the SR will be restricted to a narrative synthesis. A narrative synthesis should not simply summarize the findings from the included studies in order to draw conclusions about the body of evidence, but instead should be a more formal process which includes a formulation of the theory of how the intervention works, why and for whom, the exploration of the relationships in the data, and the assessment of the robustness of the synthesis⁵⁹.

When it is reasonable to perform a statistical synthesis, there are two main models to conduct a meta-analysis: fixed effects model and random effects model. For practical purposes, the chosen model determines how the studies included in the meta-analysis will be numerically weighed. Both models are based on different assumptions regarding distribution of effects and heterogeneity in the set of studies, and they differ in their application and interpretation⁵⁸.

Table 4. Methodological characteristics of meta-analysis by type of systematic review.

	Measures to combine	Assessment of heterogeneity	Model	Method	Command (package)
Prevalence review	- Proportion (prevalence)	- Qualitative	- Fixed/Random effects	- Inverse-variance method ^a	- Metaprop (Stata)
Prognostic review - global prognosis	- Cumulative incidence - Incidence rate	- Meta-regression	- Fixed/Random effects	- Inverse-variance method ^b	- Metan (Stata) - Metaprop (Stata) - Review Manager
Prognostic review- prognostic factors	- Hazard Ratio - Odds Ratio	- Meta-regression	- Random effects	- Inverse-variance method	- Metafor (R)
Prognostic review- prognostic models	- Calibration - Discrimination	- Meta-regression	- Random effects	- Multivariate methods	- Metamisc (R)
Diagnostic accuracy review	- Sensitivity - Specificity	- Meta-regression	- Random effects	- HSROC method ^c - Bivariate model	- Metadas (SAS) - Metandi (Stata)
Effects of intervention review	- Mean difference - Risk difference - Standardised mean difference - Hazard Ratio - Incidence rate ratio - Odds Ratio - Risk ratio	- I^2 - Meta-regression	- Fixed/Random effects	- Mantel-Haenszel method - Multivariate methods	- Metafor (R) - Metan (Stata) - Review Manager

^a Tukey-Freeman or logit transformation. ^b Transformation for the cumulative incidence. ^c Hierarchical summary receiver operating characteristic (HSROC) method allows estimation of a Receiver operating characteristic (ROC) curve or sensitivity and specificity indexes.

Finally, there is a variety of resources to conduct meta-analyses, from specific programs to perform meta-analyses (free or paid) to user-defined routines using general statistics packages (SAS, Stata, SPSS), as well as Excel utilities or R libraries. An archive with software and utilities is available from [SR Tool Box](#).

Due to the complexity of the statistical techniques to synthesise results, and the difficulty to standardise methods and decisions to be made during the analysis, it is vital to involve a statistician in the planning and conduct stages of the meta-analysis, especially for prognostic and diagnostic SRs.

Prevalence review. In prevalence SRs, the meta-analysis combines ratios, which are transformed to be meta-analysed using the inverse-variance method⁵⁸. Siriwardhana *et al.* (2018) calculated combined frailty prevalence estimates using a random effects model. The authors assessed that there was high clinical heterogeneity between the studies in terms of actual frailty prevalence, geographic setting, frailty assessment method, cut-off points applied and sample age, although this heterogeneity did not rule out performing a meta-analysis⁵.

Prognostic review. In global prognostic SRs, the meta-analysis combines cumulative incidence ratios or incidence rates, while in prognostic factor SRs, the meta-analysis combines odds ratios or hazard ratios, which can be presented in individual studies as raw estimates or as covariate-adjusted estimations derived from logistic or Cox regression models. If combining adjusted estimates, all of them should be adjusted by a minimum set of common factors¹⁷. In prognostic model SRs, the meta-analysis combines estimates of model discrimination and calibration. These indicators can be synthesised separately or jointly using multivariate models¹⁹.

Prognostic studies usually show significant variability in terms of design, sample case-mix, measurement instruments, analysis methods and presentation of results¹⁷. Therefore, in prognostic factor and model SRs, it is recommended to perform the meta-analysis using the random effects model, and even to use multivariate meta-analysis methods adjusting for relevant factors¹⁷. For instance, the SR by Westby *et al.* (2018) describes how the authors dismissed performing a meta-analysis due to the high risk of bias and the extreme heterogeneity across the included studies in terms of population, measurement of the prognostic factor (cut-off points and analytical methods) and outcome measurement⁸.

Diagnostic accuracy review. In diagnostic SRs, the meta-analysis combines estimates of sensitivity and specificity of the index test. The meta-analysis in diagnostic SRs shows a higher degree of complexity because the studies may have used different thresholds, both implicit and explicit, to define a positive result in the evaluated test. This leads to a correlation between the sensitivity and specificity indexes, which must be modelled jointly using multivariate methods⁶⁰. The most common available statistical methods are the bivariate hierarchical model and the HSROC model (Hierarchical summary receiver-operating characteristic)⁶¹. Diagnostic SRs tend to combine studies with very

heterogeneous results, and it is recommended to use the random effects model by default and perform a comprehensive examination of the sources of heterogeneity using meta-regression⁶⁰. For instance, the protocol of the SR by Ambagtsheer *et al.* (2017) expects to estimate an average sensitivity and specificity for the frailty scales, when the included studies have applied the same explicit cut-off points to the considered scales. However, given that they are subjective, self-reported scales, the studies could share the same explicit cut-off point, and yet that cut-off point could correspond to different levels of frailty in the studies (implicit thresholds), which will advise against calculating pooled estimates of diagnostic accuracy¹⁰.

Effects of interventions review. In intervention SRs, the meta-analysis combines different measures, depending on the type of outcome: odds ratio or risk ratio for binary outcomes, mean difference or standardised mean difference for continuous outcomes, hazard ratio for time-to-event outcomes, and incidence rate ratios for outcomes that count number of events.

In intervention SRs, the I^2 estimator has been proposed to assess statistical heterogeneity as a supplement to the assessment of clinical and methodological heterogeneity. This indicator is defined as the percentage of the overall variability that cannot be explained by chance, and has values ranging from 0% to 100%; with higher values indicating higher statistical heterogeneity⁵⁹. For instance, in the SR by Ellis *et al.* (2017), the authors established a 70% heterogeneity limit for I^2 , beyond which a meta-analysis combining the results would not be performed¹³. Despite its popularity and ease of interpretation, the use of this indicator is not exempt of controversy due to its dependence on the number of studies and sample size; thus, a small statistical heterogeneity could seem substantial only by the effect of a large sample size of the included studies⁶².

In intervention SRs, pairwise meta-analysis has been extended to network meta-analysis, which allows the simultaneous comparison of three or more interventions, combining direct and indirect evidence from a network of studies⁶³.

Quality of evidence

The quality (also confidence or certainty) of evidence in an SR is the degree of confidence that is held against the fact that an estimate of effect or association is close to the actual value of interest¹. Certainty of evidence is best evaluated with the GRADE system. Certainty in the obtained estimates for each one of the key SR outcomes or factors is classified as high, moderate, low or very low. A level of certainty of evidence is first established from the design of the studies that form the evidence body, which might or might not have an optimal design for the type of considered question. This initial confidence in the evidence body can then decrease in one or two levels if the following is detected: 1) design or execution limitations, 2) inconsistency between estimates, 3) indirect evidence, 4) imprecision in estimates, or 5) publication bias⁶⁴.

The certainty of evidence is a key element to interpret and communicate results, and as such, it should be included in the sections of results, discussion, conclusion and abstract, using

semi-standardised statements⁶⁵. Additionally, it can be included in a Summary of Findings table, where for each comparison, the key information regarding relative effect and absolute effect magnitude, quantity of available evidence and its certainty is presented⁶⁶. Certainty of evidence can be assessed too when no quantitative synthesis is possible⁶⁷.

We will now highlight the specific aspects in which the GRADE system adapts to each type of SR.

Prevalence review. There are no formal adaptations of the GRADE system for prevalence SRs, but there is a proposal to assess the quality of the evidence based on this system⁶⁸. High initial certainty is awarded to survey or cross-sectional study designs with population representativeness that have been properly designed and conducted, while studies with no population representativeness will have lower initial quality.

Prognostic review. There is a GRADE proposal for global prognostic SRs²² and an adaptation for prognostic factor SR⁶⁹. Guidelines for prognostic model SR are still under development.

In **global prognostic** SRs, the study designs that have high initial certainty are longitudinal cohort studies and pragmatic randomised controlled trials with representative samples²². Other observational designs would offer low initial certainty. In **prognostic factor** SRs, explanatory and confirmatory longitudinal designs offer high initial certainty, while exploratory studies are considered to be of moderate quality⁶⁹.

In prognostic SRs, the assessment of the limitations follows the general procedure already described, with two particularities: 1) qualitative assessment of inconsistency, because of low reliability of I^2 estimator in the prognostic field^{22,69}, and 2) possibility of increased certainty in the studies that do not show limitations in the quality of evidence, if (i) the estimated effect magnitude is substantial, or (ii) there is an exposure-response gradient⁶⁷. For instance, the prognostic factor SR by Westby *et al.* (2018) considered the possibility of increasing the certainty of evidence in the studies presenting no limitations. Due to the exploratory nature of the included studies and their high risk of bias, the certainty was not increased in any case and the evidence obtained in the review was of very low quality⁸.

Diagnostic accuracy review. There is a GRADE proposal for assessing the certainty of evidence for test accuracy^{70,71}. The study designs that start with the highest degree of evidence are cohort or cross-sectional studies where the index test and an appropriate reference standard have been directly compared in patients with diagnostic uncertainty⁷⁰. If the SR included case-control studies, these would offer low-quality initial evidence²⁵.

Indirectness of evidence would be assessed through any applicability concerns of the patient sample, the intervention and the comparator with respect to the clinical pathway where the test is to be applied. There is uncertainty regarding how to assess

inconsistency, because heterogeneity is common and hard to quantify in diagnostic SRs, and it often cannot be explained even if multivariate models are adjusted. Judgments on extent of heterogeneity should be based on similarity of the point estimates, overlap of confidence intervals, and the exploration of possible explanations for the inconsistency from subgroup or sensitivity analyses⁷⁰.

Imprecision judgments should be based on both the width of the confidence or credible intervals for sensitivity and specificity, but also on the implications for patient management in terms of true and false positives, and true and false negatives. When the estimated intervals include values that may lead to different conclusions of the test's value, the certainty of the evidence may be lowered⁷¹.

With regard to the criteria to increase the level of evidence, it is unclear whether they should be applied at all and how to do it in diagnostic SRs⁷¹. The uncertainty surrounding the process of assessing the quality of evidence in diagnostic SRs explains why it is not a requirement in Cochrane SRs at the moment. For instance, the SR by Martínez *et al.* (2017) only included a Summary of Findings table with numerical results and an estimation of the absolute effect that the test would have on a hypothetical cohort of individuals¹¹.

Effects of interventions review. The GRADE system for assessing the quality of evidence was initially developed for intervention SRs, and it is the indication for which clearer and widely agreed guidelines are available⁶⁴. In terms of study design, RCTs are initially classified as having high certainty, while all non-randomised or observational studies are classified as having low certainty. This proposal for pairwise metaanalysis can be extended to network metaanalysis⁶³.

The assessment of the certainty limitations is well-defined in intervention SRs. Inconsistency can be assessed using the I^2 estimator⁶⁴. Imprecision is assessed taking into account whether the review meets the optimal information size, and whether the confidence interval of the effect estimate allows reaching a conclusion, because either it only includes values consistent with a relevant intervention effect, or it completely dismisses it⁶³. In observational studies that do not have limitations in the quality of evidence, three criteria are considered to increase certainty: 1) the estimated effect magnitude is important or very important, 2) there is an exposure-response gradient, and 3) all possible biases that could reduce the observed effect confirm the obtained conclusions.

For instance, the SR by Ellis *et al.* (2017) applied the GRADE system to the included randomised trials, and it concluded that there was high certainty of the effect of the comprehensive geriatric assessment on the effects outcomes based on a high number of studies and participants, with a globally low risk of bias, and results consistent among studies. However, the certainty of evidence obtained in cost-effectiveness was low, due to imprecision and inconsistency of results¹³.

Results report

It is vital to inform about the methods, results and conclusions of the SRs in a transparent and thorough manner so that their users can interpret, evaluate and apply them. The **EQUATOR initiative** has developed, and keeps up-to-date, a library with guidelines to communicate the different types of research studies. The **PRISMA statement** (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) has been proposed in the SR field⁷². This statement consists of a checklist comprised of 27 items and a flow diagram to present the number of studies considered in the SR. In addition, several extensions focusing on reporting specific aspects of SRs have been developed, such as PRISMA-P for reporting SR protocols⁷³, PRISMA- Abstracts for reporting abstracts⁷⁴, and PRISMA- Harms for reporting harms outcomes in SRs⁷⁵. Additionally, the SWiM guideline is available for reporting intervention SRs where the effects of interventions are synthesised narratively without metaanalysis, focusing on the key features of narrative information synthesis (grouping of studies, presentation of data and summary text, and appropriate discussion of limitations of this type of synthesis)⁷⁶.

Although the PRISMA statement and the cited extensions are focused on intervention SRs, a specific PRISMA extension has also been developed for diagnostic SRs⁷⁷. On the contrary, no tools have been identified to communicate prevalence or prognostic SRs. In recent years, clarity and transparency in study communications has improved thanks to the development of checklists for scientific paper publication, although there is still room for improvement^{78–80}.

Discussion

Key results

This review identifies and describes the most relevant methodological resources to conduct prevalence, prognostic, diagnostic accuracy and effects of interventions SRs. This review offers a general and comparative perspective of the methodological resources by SR stage, highlighting the differential elements of each type of SR. This project does not aim to be a standalone tool for a researcher to find complete guidance on how to conduct and report a review, but rather it aims to be a signpost pointing out to the resources where researchers may find in depth guidance to develop their reviews.

Current context

This paper corroborates that developing a rigorous SR is a complex and resource-intensive task^{81,82}. In order to tackle the increasing complexity of SRs and ensure the adoption of rigorous methodology, it is necessary that the reviews are made by **multi-disciplinary work groups** with knowledge and experience in methodology (such as statistical analysis and information retrieval)^{83,84}. In addition, it is important to consider the increasing availability of artificial-intelligence-based **technological tools**, which make it possible to semi-automate the different steps of the SR development, and thus reduce the time and human resources required to conduct the review⁸⁵.

Once the rigorous SR has been developed, ensuring the conveyance of the generated knowledge is essential. In this sense,

new **formats for synthesis** and presentation of SR results are being explored nowadays to help their dissemination and the adoption of their conclusions in clinical practice and healthcare decision-making. For instance, new formats for result presentation and Summary of Findings tables are being proposed, adapted to the profile of their potential users^{86,87}.

Limitations and strengths

An inherent limitation of this project is its methodology based on a selection of resources and summary of guidance informed by expert opinion, which may be susceptible to implicit selection biases or lack of comprehensiveness.

The four types of SRs considered in this paper are fundamental to define preventive activities and public health policies, as well as to make health decisions. The selection of resources done is not dependent on whether the reviewer explores questions on efficacy or effectiveness, often described as explanatory or pragmatic questions, and will be useful to the researchers regardless of their intended purpose. However, we have not considered the resources to conduct in-depth exploration of effectiveness issues such as reviews of complex interventions or implementation reviews. Additionally, this research has not considered other types of SRs, such as methodological, economic evaluation and qualitative research SRs, for which it would be convenient to perform similar methodological compilations. Reviews of reviews (or overviews) were also not considered, and as such, there are a number of review-level resources which have not been discussed, for example the risk of bias assessment tool ROBIS or the methodological assessment tool AMSTAR^{88,89}. Another limitation of this research is the need to keep it up to date, given the speed at which the methods and methodological resources to develop SRs are updated.

On the other hand, the main strengths of this paper are its transversal approach for the different types of reviews, and the identification of resources for all the stages in the development of an SR. There are few previous publications that offer a transversal perspective of the different types of systematic reviews, and these are focused on a specific stage of the review or on a particular topic. For instance, the work carried out by Munn *et al.* (2018) defined a typology for SRs, characterised from 10 different types of research questions, and delving into the format of each type of question²¹. Pollock *et al.* (2017) review the steps of an SR for 5 types of question, specifically focusing on the particularities of the reviews on stroke rehabilitation⁹⁰. Muka *et al.* (2019) offer a structured compilation of resources for each SR stage, but without delving into the specificities of the different types of SRs⁹¹. Finally, organising the resources to assess the risk of bias by type of review is a strength and a novelty compared with previous works, which compile the quality assessing tools by type of study design but without linking them to the aim of the study nor the type of systematic review^{92,93}.

Conclusions

SRs are a key research tool to make decisions in healthcare, public health and medical research. There are methods and resources to develop high-quality reviews to answer most types of clinical questions. This review offers a complete resource guide for prevalence, prognostic, diagnostic and intervention reviews,

and is a very useful tool for those researchers that wish to develop SRs or conduct methodological research works in that field.

Data availability

Underlying data

All data underlying the results are available as part of the article and no additional source data are required.

Acknowledgements

Marta Roqué i Figuls is presently working on her PhD with the PhD Programme in Biomedical Research Methodology and Public Health from the Autonomous University of Barcelona.

The authors are indebted to the helpful peer reviewer comments of Dr Cumpston and Dr Purcell, which have helped improve the manuscript.

References

- Higgins JPT, Thomas J, Chandler J, *et al.*: (editors). **Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019)**. Cochrane. 2019. [Accessed on 29/11/2019].
[Reference Source](#)
- Urrútia G, Bonfill X: **Revisión sistemática, una herramienta clave para la toma de decisiones clínicas y sanitarias**. *Rev Esp Salud Pública*. 2014; **88**(1): 1–3.
[Publisher Full Text](#)
- Ferreira González I, Urrútia G, Alonso-Coello P: **Systematic reviews and meta-analysis: scientific rationale and interpretation**. *Rev Esp Cardiol*. 2011; **64**(8): 688–96.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hemingway H, Croft P, Perel P, *et al.*: **Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes**. *BMJ*. 2013; **346**: e5595.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Siriwardhana DD, Hardoon S, Rait G, *et al.*: **Prevalence of Frailty and Prefrailty Among Community-Dwelling Older Adults in Low-Income and Middle-Income Countries: A Systematic Review and Meta-Analysis**. *BMJ Open*. 2018; **8**(3): e018195.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Guthold R, Stevens GA, Riley LM, *et al.*: **Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants**. *Lancet Glob Health*. 2018; **6**(10): e1077–86.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Roehr S, Pabst A, Luck T, *et al.*: **Is dementia incidence declining in high-income countries? A systematic review and meta-analysis**. *Clin Epidemiol*. 2018; **10**: 1233–1247.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Westby MJ, Dumville JC, Stubbs N, *et al.*: **Protease activity as a prognostic factor for wound healing in venous leg ulcers**. *Cochrane Database Syst Rev*. 2018; **9**(9): CD012841.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Skoetz N, Trivella M, Kreuzer KA, *et al.*: **Prognostic models for chronic lymphocytic leukaemia: an exemplar systematic review and meta-analysis**. *Cochrane Database of Syst Rev*. 2016; **1**: CD012022.
[Publisher Full Text](#)
- Ambagtsheer RC, Thompson MQ, Archibald MM, *et al.*: **Diagnostic test accuracy of self-reported frailty screening instruments in identifying community-dwelling older people at risk of frailty and pre-frailty: a systematic review protocol**. *JBI Database System Rev Implement Rep*. THE JOANNA BRIGGS INSTITUTE. 2017; **15**(10): 2464–2468.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Martínez G, Vernooij RW, Fuentes Padilla P, *et al.*: **18F PET with florbetaben for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)**. *Cochrane Database Syst Rev*. 2017; **11**(11): CD012883.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Johnson S, Henschke N, Maayan N, *et al.*: **Ribavirin for treating Crimean Congo haemorrhagic fever**. *Cochrane Database Syst Rev*. 2018; **6**(6): CD012713.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ellis G, Gardner M, Tsiachristas A, *et al.*: **Comprehensive geriatric assessment for older adults admitted to hospital**. *Cochrane Database Syst Rev*. 2017; **9**(1): CD006211.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Aromataris E, Munn Z, (Editors): **Joanna Briggs Institute Reviewer's Manual**. The Joanna Briggs Institute, 2017. [Accessed on 10/12/2018].
- Munn Z, Moola S, Lisy K, *et al.*: **Chapter 5: Systematic reviews of prevalence and incidence**. In: Aromataris E, Munn Z, (Editors). *Joanna Briggs Institute Reviewer's Manual*. The Joanna Briggs Institute. 2017. [Accessed on 10/12/2018].
[Reference Source](#)
- Moola S, Munn Z, Tufanaru C, *et al.*: **Chapter 7: Systematic reviews of etiology and risk**. In: Aromataris E, Munn Z, (Editors). *Joanna Briggs Institute Reviewer's Manual*. The Joanna Briggs Institute. 2017. [Accessed on 10/12/2018].
[Reference Source](#)
- Riley RD, Moons KGM, Snell KIE, *et al.*: **A guide to systematic review and meta-analysis of prognostic factor studies**. *BMJ*. 2019; **364**: k4597.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dekkers OM, Vandenbroucke JP, Cevallos M, *et al.*: **COSMOS-E: Guidance on conducting systematic reviews and meta-analyses of observational studies of etiology**. *PLoS Med*. 2019; **16**(2): e1002742.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Debray TP, Damen JA, Snell KI, *et al.*: **A guide to systematic review and meta-analysis of prediction model performance**. *BMJ*. 2017; **356**: i6460.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Deeks JJ, Bossuyt PM, Gatsonis C, (editors): **Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0**. The Cochrane Collaboration. 2010; [Accessed on 21/12/2018].
[Reference Source](#)
- Munn Z, Stern C, Aromataris E, *et al.*: **What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences**. *BMC Med Res Methodol*. 2018; **18**(1): 5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Iorio A, Spencer FA, Falavigna M, *et al.*: **Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients**. *BMJ*. 2015; **350**: h870.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bossuyt PM, Irwig L, Craig J, *et al.*: **Comparative accuracy: assessing new tests against existing diagnostic pathways**. *BMJ*. 2006; **332**(7549): 1089–92. Erratum in: *BMJ*. 2006 Jun 10; **332**(7554): 1368.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bossuyt PM, Leflang MM: **Chapter 6: Developing Criteria for Including Studies**. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4* [updated September 2008]. The Cochrane Collaboration. 2008.
[Reference Source](#)
- Lijmer JG, Mol BW, Heisterkamp S, *et al.*: **Empirical evidence of design-related bias in studies of diagnostic tests**. *JAMA*. 1999; **282**(11): 1061–6.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Straus S, Moher D: **Registering systematic reviews**. *CMAJ*. 2010; **182**(1): 13–14.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ge L, Tian JH, Li YN, *et al.*: **Association between prospective registration and overall reporting and methodological quality of systematic reviews: a meta-epidemiological study**. *J Clin Epidemiol*. 2018; **93**: 45–55.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Page MJ, Shamseer L, Tricco AC: **Registration of systematic reviews in PROSPERO: 30,000 records and counting**. *Syst Rev*. 2018; **7**(1): 32.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lefebvre C, Glanville J, Briscoe S, *et al.*: **Chapter 4: Searching for and selecting studies**. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.0*. (updated July 2019). Cochrane, 2019; [Accessed on 29/11/2018].
[Publisher Full Text](#)
- Atkinson KM, Koenka AC, Sanchez CE, *et al.*: **Reporting standards for literature searches and report inclusion criteria: making research syntheses more transparent and easy to replicate**. *Res Synth Methods*. 2015; **6**(1): 87–95.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Campbell JM, Kulgar M, Ding S, *et al.*: **Chapter 9: Diagnostic test accuracy systematic reviews**. In: Aromataris E, Munn Z, (Editors). *Joanna Briggs Institute Reviewer's Manual*. The Joanna Briggs Institute, 2017; [Accessed on 10/12/2018].
[Reference Source](#)
- de Vet HCW, Eisinga A, Riphagen II, *et al.*: **Chapter 7: Searching for Studies**. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4* [updated September 2008]. The Cochrane Collaboration, 2008.
[Reference Source](#)

33. Lefebvre C, Glanville J, Wieland LS, *et al.*: **Methodological developments in searching for studies for systematic reviews: past, present and future?** *Syst Rev.* 2013; 2: 78.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Glanville JM, Lefebvre C, Miles JN, *et al.*: **How to identify randomized controlled trials in MEDLINE: ten years on.** *J Med Libr Assoc.* 2006; 94(2): 130–136.
[PubMed Abstract](#) | [Free Full Text](#)
35. Wilczynski NL, Haynes RB, Hedges Team: **Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey.** *BMC Med.* 2004; 2(1): 23.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Beynon R, Leeflang MM, McDonald S, *et al.*: **Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE.** *Cochrane Database Syst Rev.* 2013; (9): MR000022.
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Sampson M, Tetzlaff J, Urquhart C: **Precision of healthcare systematic review searches in a cross-sectional sample.** *Res Synth Methods.* 2011; 2(2): 119–25.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Bramer WM, Rethlefsen ML, Kleijnen J, *et al.*: **Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study.** *Syst Rev.* 2017; 6(1): 245.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Hartling L, Featherstone R, Nussli M, *et al.*: **The contribution of databases to the results of systematic reviews: a cross-sectional study.** *BMC Med Res Methodol.* 2016; 16(1): 127.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Glanville JM, Duffy S, McCool R, *et al.*: **Searching ClinicalTrials.gov and the International Clinical Trials Registry Platform to inform systematic reviews: what are the optimal search approaches?** *J Med Libr Assoc.* 2014; 102(3): 177–83. MANUAL GRADE. Spanish version. [Accessed May 2019].
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
41. Isojarvi J, Wood H, Lefebvre C, *et al.*: **Challenges of identifying unpublished data from clinical trials: Getting the best out of clinical trials registers and other novel sources.** *Res Synth Methods.* 2018; 9(4): 561–578.
[PubMed Abstract](#) | [Publisher Full Text](#)
42. Horsley T, Dingwall O, Sampson M: **Checking reference lists to find additional studies for systematic reviews.** *Cochrane Database Syst Rev.* 2011; (8): MR000026.
[PubMed Abstract](#) | [Publisher Full Text](#)
43. Gentles SJ, Charles C, Nicholas DB, *et al.*: **Reviewing the research methods literature: principles and strategies illustrated by a systematic overview of sampling in qualitative research.** *Syst Rev.* 2016; 5(1): 172.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. Hartling L, Featherstone R, Nussli M, *et al.*: **Grey literature in systematic reviews: a cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews.** *BMC Med Res Methodol.* 2017; 17(1): 64.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. Booth A: **Searching for qualitative research for inclusion in systematic reviews: a structured methodological review.** *Syst Rev.* 2016; 5: 74.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
46. Rethlefsen ML, Murad MH, Livingston EH: **Engaging medical librarians to improve the quality of review articles.** *JAMA.* 2014; 312(10): 999–1000.
[PubMed Abstract](#) | [Publisher Full Text](#)
47. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, *et al.*: **Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews.** *J Clin Epidemiol.* 2015; 68(6): 617–26.
[PubMed Abstract](#) | [Publisher Full Text](#)
48. Spencer AJ, Eldredge JD: **Roles for librarians in systematic reviews: a scoping review.** *J Med Libr Assoc.* 2018; 106(1): 46–56.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. Hoy D, Brooks P, Woolf A, *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.
[PubMed Abstract](#) | [Publisher Full Text](#)
50. 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.
[PubMed Abstract](#) | [Publisher Full Text](#)
51. Morgan RL, Thayer KA, Santesso N, *et al.*: **Evaluation of the risk of bias in non-randomized studies of interventions (ROBINS-I) and the 'target experiment' concept in studies of exposures: Rationale and preliminary instrument development.** *Environ Int.* 2018; 120: 382–387.
[PubMed Abstract](#) | [Publisher Full Text](#)
52. Morgan RL, Thayer KA, Santesso N, *et al.*: **A risk of bias instrument for non-randomized studies of exposures: A users' guide to its application in the context of GRADE.** *Environ Int.* 2019; 122: 168–184.
[PubMed Abstract](#) | [Publisher Full Text](#)
53. Wolff RF, Moons KGM, Riley RD, *et al.*: **PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies.** *Ann Intern Med.* 2019; 170(1): 51–58.
[PubMed Abstract](#) | [Publisher Full Text](#)
54. Whiting PF, Rutjes AW, Westwood ME, *et al.*: **QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies.** *Ann Intern Med.* 2011; 155(8): 529–36.
[PubMed Abstract](#) | [Publisher Full Text](#)
55. Higgins JPT, Savovi J, Page MJ, *et al.*: **Chapter 8: Assessing risk of bias in a randomized trial.** In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions.* version 6.0 (updated July 2019). Cochrane, 2019.
[Publisher Full Text](#)
56. Sterne JA, Hernán MA, Reeves BC, *et al.*: **ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions.** *BMJ.* 2016; 355: i4919.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
57. Lau J, Ioannidis JP, Schmid CH: **Quantitative synthesis in systematic reviews.** *Ann Intern Med.* 1997; 127(9): 820–826.
[PubMed Abstract](#) | [Publisher Full Text](#)
58. Deeks JJ, Higgins JPT, Altman DG: (editors). **Chapter 10: Analysing data and undertaking meta-analyses.** In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.0* (updated July 2019). Cochrane, 2019.
[Reference Source](#)
59. Popay J, Roberts H, Sowden A, *et al.*: **Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme.** 2006.
[Publisher Full Text](#)
60. Macaskill P, Gatsonis C, Deeks JJ, *et al.*: **Chapter 10: Analysing and Presenting Results.** In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.* The Cochrane Collaboration, 2010.
[Reference Source](#)
61. Rutter CM, Gatsonis CA: **A Hierarchical Regression Approach to Meta-Analysis of Diagnostic Test Accuracy Evaluations.** *Stat Med.* 2001; 20(19): 2865–84.
[PubMed Abstract](#) | [Publisher Full Text](#)
62. Rücker G, Schwarzer G, Carpenter JR, *et al.*: **Undue Reliance on I^2 in Assessing Heterogeneity May Mislead.** *BMC Med Res Methodol.* 2008; 8(1): 79.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
63. Chaimani A, Caldwell DM, Higgins T, *et al.*: **Chapter 11: Undertaking network meta-analyses.** In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions.* version 6.0 (updated July 2019). Cochrane, 2019.
[Publisher Full Text](#)
64. Schünemann H, Broek J, Guyatt G, *et al.*: (editors). **Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach.** (updated October 2013). [Accessed June 2020].
[Reference Source](#)
65. Santesso N, Glenton C, Dahm P, *et al.*: **GRADE Guidelines 26: Informative Statements to Communicate the Findings of Systematic Reviews of Interventions.** *J Clin Epidemiol.* 2020; 119: 126–135.
[PubMed Abstract](#) | [Publisher Full Text](#)
66. Schünemann HJ, Higgins JPT, Vist GE, *et al.*: **Chapter 14: Completing 'Summary of findings' tables and grading the certainty of the evidence.** In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions version 6.0* (updated July 2019). Cochrane, 2019.
[Reference Source](#)
67. Murad MH, Mustafa RA, Schünemann HJ, *et al.*: **Rating the certainty in evidence in the absence of a single estimate of effect.** *Evid Based Med.* 2017; 22(3): 85–87.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
68. Harder T, Takla A, Eckmanns T, *et al.*: **PRECEPT: An Evidence Assessment Framework for Infectious Disease Epidemiology, Prevention and Control.** *Euro Surveill.* 2017; 22(40): 16-00620.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
69. Huet A, Hayden JA, Stinson J, *et al.*: **Judging the Quality of Evidence in Reviews of Prognostic Factor Research: Adapting the GRADE Framework.** *Syst Rev.* 2013; 2: 71.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
70. Schünemann HJ, Mustafa RA, Brozek J, *et al.*: **GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of evidence for test accuracy.** *J Clin Epidemiol.* 2020; 122: 129–141.
[PubMed Abstract](#) | [Publisher Full Text](#)
71. Schünemann HJ, Mustafa RA, Brozek J, *et al.*: **GRADE guidelines: 21 part 2. Test accuracy: inconsistency, imprecision, publication bias, and other domains for rating the certainty of evidence and presenting it in evidence profiles and summary of findings tables.** *J Clin Epidemiol.* 2020; 122: 142–152.
[PubMed Abstract](#) | [Publisher Full Text](#)
72. Moher D, Liberati A, Tetzlaff J, *et al.*: **Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.** *BMJ.* 2009; 339: b2535.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
73. Moher D, Shamseer L, Clarke M, *et al.*: **Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 Statement.** *Syst Rev.* 2015; 4(1): 1.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
74. Beller EM, Glasziou PP, Altman DG, *et al.*: **PRISMA for Abstracts: Reporting**

- Systematic Reviews in Journal and Conference Abstracts.** *PLoS Med.* 2013; 10(4): e1001419.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
75. Zorzela L, Loke YK, Ioannidis JP, *et al.*: **PRISMA Harms Checklist: Improving Harms Reporting in Systematic Reviews.** *BMJ.* 2016; 352: i157.
[PubMed Abstract](#) | [Publisher Full Text](#)
 76. Campbell M, McKenzie JE, Sowden A, *et al.*: **Synthesis without meta-analysis (SWIM) in systematic reviews: reporting guideline.** *BMJ.* 2020; 368: l6890.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 77. McInnes MDF, Moher D, Thoms BD, *et al.*: **Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement.** *JAMA.* 2018; 319(4): 388–396.
[PubMed Abstract](#) | [Publisher Full Text](#)
 78. Moher D, Tetzlaff J, Tricco AC, *et al.*: **Epidemiology and Reporting Characteristics of Systematic Reviews.** *PLoS Med.* 2007; 4(3): e78.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 79. 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.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 80. Salameh JP, McInnes MDF, Moher D, *et al.*: **Completeness of Reporting of Systematic Reviews of Diagnostic Test Accuracy Based on the PRISMA-DTA Reporting Guideline.** *Clin Chem.* 2019; 65(2): 291–301.
[PubMed Abstract](#) | [Publisher Full Text](#)
 81. Turner T, Green S, Tovey D, *et al.*: **Producing Cochrane systematic reviews-a qualitative study of current approaches and opportunities for innovation and improvement.** *Syst Rev.* 2017; 6(1): 147.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 82. Borah R, Brown AW, Capers PL, *et al.*: **Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry.** *BMJ Open.* 2017; 7(2): e012545.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 83. Ioannidis JP, Greenland S, Hlatky MA, *et al.*: **Increasing Value and Reducing Waste in Research Design, Conduct, and Analysis.** *Lancet.* 2014; 383(9912): 166–75.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 84. Institute of Medicine (US) Committee on Standards for Systematic Reviews of Comparative Effectiveness Research, Eden J, Levit L, *et al.*: **Finding What Works in Health Care: Standards for Systematic Reviews.** Washington (DC): National Academies Press (US); 2011.2, Standards for Initiating a Systematic Review. 2011.
[PubMed Abstract](#) | [Publisher Full Text](#)
 85. Marshall IJ, Wallace BC: **Toward systematic review automation: a practical guide to using machine learning tools in research synthesis.** *Syst Rev.* 2019; 8(1): 163.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 86. Carrasco-Labra A, Brignardello-Petersen R, Santesso N, *et al.*: **Improving GRADE evidence tables part 1: a randomized trial shows improved understanding of content in summary of findings tables with a new format.** *J Clin Epidemiol.* 2016; 74: 7–18.
[PubMed Abstract](#) | [Publisher Full Text](#)
 87. Marquez C, Johnson AM, Jassemi S, *et al.*: **Enhancing the uptake of systematic reviews of effects: what is the best format for health care managers and policy-makers? A mixed-methods study.** *Implement Sci.* 2018; 13(1): 84.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 88. Whiting P, Savovi J, Higgins JP, *et al.*: **ROBIS: A new tool to assess risk of bias in systematic reviews was developed.** *J Clin Epidemiol.* 2016; 69: 225–234.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 89. Shea BJ, Reeves BC, Wells G, *et al.*: **AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both.** *BMJ.* 2017; 358: j4008.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 90. Pollock A, Berge E: **How to do a systematic review.** *Int J Stroke.* 2018; 13(2): 138–56.
[PubMed Abstract](#) | [Publisher Full Text](#)
 91. Muka T, Glisic M, Milic J, *et al.*: **A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research.** *Eur J Epidemiol.* 2020; 35(1): 49–60.
[PubMed Abstract](#) | [Publisher Full Text](#)
 92. Harrison JK, Reid J, Quinn TJ, *et al.*: **Using Quality Assessment Tools to Critically Appraise Ageing Research: A Guide for Clinicians.** *Age Ageing.* 2017; 46(3): 359–65.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 93. Zeng X, Zhang Y, Kwong JS, *et al.*: **The Methodological Quality Assessment Tools for Preclinical and Clinical Studies, Systematic Review and Meta-Analysis, and Clinical Practice Guideline: A Systematic Review.** *J Evid Based Med.* 2015; 8(1): 2–10.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 07 October 2020

<https://doi.org/10.5256/f1000research.28216.r69215>

© 2020 Purssell E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Edward Purssell 

School of Health Sciences, City, University of London, London, UK

I would just push point 5 regarding heterogeneity. Cochrane is quite clear about this - stating "Thus, the test for heterogeneity is irrelevant to the choice of analysis; heterogeneity will always exist whether or not we happen to be able to detect it using a statistical test"¹. I think that this just needs a bit more explanation.

References

1. Higgins J, Thomas J, *et al.*: Cochrane Handbook for Systematic Reviews of Interventions (Version 6.1). 2020. [Reference Source](#)

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Paediatrics, infection control, systematic review.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 01 September 2020

<https://doi.org/10.5256/f1000research.28216.r69216>

© 2020 Cumpston M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Miranda Cumpston 

¹ Monash University, Melbourne, Australia

² University of Newcastle, Newcastle, Australia

Thanks to the authors for their work in revising this paper and responding effectively to the comments raised in peer review. I have no further comments to raise.

Thanks also for including me in the Acknowledgements - please note that my correct title is 'Ms'.

Competing Interests: I am an Associate Editor of the Cochrane Handbook for Systematic Reviews, and an Editor with Cochrane Public Health.

Reviewer Expertise: In presenting my comments, I acknowledge that my own experience focuses on methods around systematic reviews of the effects of interventions, and so I cannot comment with expertise on the details of methods described for other review types.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 28 February 2020

<https://doi.org/10.5256/f1000research.24297.r59540>

© 2020 Purssell E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Edward Purssell 

School of Health Sciences, City, University of London, London, UK

I have a few suggestions which may be helpful to the authors:

1. When discussing research methods, it may be worth just mentioning the efficacy/effectiveness issue.
2. I think that this "To obtain an efficient search with adequate sensitivity, performing searches in [MEDLINE](#) and [EMBASE](#) is sufficient, as they are the two most frequently used bibliographic databases³⁹, and they are enough to identify most relevant studies for a specific SR⁴⁰" might be a bit of a sweeping statement - are the two databases always sufficient? I think it would be helpful to mention some others as well - unless the authors really do believe that these are enough.
3. "involving a medical librarian can be convenient to improve the search quality⁴⁷⁻⁴⁹". I am not sure that convenient is quite the right word.
4. Be clear to differentiate risk of bias at the study level from RoB at the review level. For example RoB2 is study-level, ROBIS is review level.

5. "For instance, in the SR by Ellis *et al.* (2017), the authors established a 70% heterogeneity limit for I^2 , beyond which a meta-analysis combining the results would not be performed¹³" - why 70%? This does not sound like a sensible decision making process anyway. Either the authors thought it was worth doing a MA in which case the heterogeneity form parts of the results, or it is not worth doing in which case this would not be a consideration. Also not doing a MA may have the effect of hiding this heterogeneity. There are also ways of investigating heterogeneity that can be illuminating.
6. Mention ROBIS and AMSAR2 as review-level tools.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

No

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Paediatrics, infection control, systematic review.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 19 Jul 2020

Marta Roqué, Iberoamerican Cochrane Centre - Sant Pau Biomedical Research Institute (IIB-Sant Pau), Barcelona, Spain

- When discussing research methods, it may be worth just mentioning the efficacy/effectiveness issue.

ANSWER: We agree that this is an important issue, which merits a full discussion that unfortunately falls outside the scope of this project. The selection of resources done is not dependent on whether the reviewer explores efficacy or effectiveness, and will be useful to the researchers regardless of their intended purpose. However, we have clarified this issue

throughout the text by referring to 'effects of interventions' (rather than efficacy of interventions) and we have explicitly commented on it in the discussion. We have substituted 'efficacy of interventions' by 'effects of interventions'. We have added the following limitation "The selection of resources done is not dependent on whether the reviewer explores questions on efficacy or effectiveness, which are often described as explanatory or pragmatic questions, and will be useful to the researchers regardless of their intended purpose. However, we have not considered the resources to conduct in-depth exploration of effectiveness issues such as reviews of complex interventions or implementation reviews."

- I think that this "To obtain an efficient search with adequate sensitivity, performing searches in MEDLINE and EMBASE is sufficient, as they are the two most frequently used bibliographic databases³⁹, and they are enough to identify most relevant studies for a specific SR⁴⁰" might be a bit of a sweeping statement - are the two databases always sufficient? I think it would be helpful to mention some others as well - unless the authors really do believe that these are enough.

ANSWER: Certainly, while MEDLINE and EMBASE are most used databases, there are other databases which can provide complementary information. However, the bibliographic references provided in the text supports the notion that these two databases are enough for an efficient search regardless of topic, and the role of these other databases is mostly complementary, but with little added benefit. To avoid the implicit suggestion that specialized databases cannot generate any added value, a sentence has been added mentioning them. Two sentences have been added "To obtain an efficient search with adequate sensitivity, performing searches in MEDLINE and EMBASE is may be sufficient especially in intervention reviews". " These searches can be complemented with additional searches in other databases such as PEDro (Hyperlink: <https://www.pedro.org.au/>), which may provide specific information for certain topics. "

- "involving a medical librarian can be convenient to improve the search quality⁴⁷⁻⁴⁹". I am not sure that convenient is quite the right word.

ANSWER: Thanks, we've substituted the term 'convenient' for 'desirable'

- Be clear to differentiate risk of bias at the study level from RoB at the review level. For example RoB2 is study-level, ROBIS is review level.

ANSWER: Thanks for pointing out this issue, we have clarified in the risk of bias section that this project focuses in presenting study-level resources, and that tools for reviews are not included in the manuscript. As we expand later in comment 6, we have been more explicit about the exclusion of overviews in the eligibility criteria section, and we have added a mention to this fact in the discussion section, as a limitation of the review, with an explicit mention of these tools. A sentence has been added "Assessing the risk of bias of the included studies is a key element in any SR". We have modified the eligibility criteria ("We excluded the methodological resources to develop other types of SRs (methodological, economic evaluation and qualitative research SRs, or overviews)" and the discussion ("Reviews of reviews (or overviews) were also not considered, and as such, there are a number of review-level resources which have not been discussed, for example the risk of bias assessment tool ROBIS or the methodological tool AMSTAR2")

- "For instance, in the SR by Ellis et al. (2017), the authors established a 70% heterogeneity limit for I², beyond which a meta-analysis combining the results would not be performed¹³" - why 70%? This does not sound like a sensible decision making process anyway. Either the authors thought it was worth doing a MA in which case

the heterogeneity form parts of the results, or it is not worth doing in which case this would not be a consideration. Also not doing a MA may have the effect of hiding this heterogeneity. There are also ways of investigating heterogeneity that can be illuminating.

ANSWER: We agree with the peer reviewer that the analysis choices of the examples shown may not always be universally shared. We chose the examples based on several considerations, as they had to be useful for the different parts of the manuscript, but they are not necessarily a perfect role model in all their decisions. Readers will need to take into account that these examples are only illustrative, as we are presenting these choices descontextualized, without a full vision of the authors decision making process.

- o Mention ROBIS and AMSTAR2 as review-level tools.

ANSWER: Thanks for this suggestion. We restricted our focus to a limited list of SR types, not including overviews (or reviews of reviews). For this reason, only study-level resources for assessing risk of bias have been described, and review-level resources (such as ROBIS or AMSTAR2) are not mentioned. We realize that this issue may not be clear enough, and consequently we have been more explicit about the exclusion of overviews in the eligibility criteria section, and we have added a mention to this fact in the discussion section, as a limitation of the review, with an explicit mention of these tools. We have modified the eligibility criteria ("We excluded the methodological resources to develop other types of SRs (methodological, economic evaluation and qualitative research SRs, or overviews)" and the discussion ("Reviews of reviews (or overviews) were also not considered, and as such, there are a number of review-level resources which have not been discussed, for example the risk of bias assessment tool ROBIS or the methodological tool AMSTAR2")

Competing Interests: No competing interests were disclosed.

Reviewer Report 14 February 2020

<https://doi.org/10.5256/f1000research.24297.r59538>

© 2020 Cumpston M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Miranda Cumpston

¹ Monash University, Melbourne, Australia

² University of Newcastle, Newcastle, Australia

The authors have summarised the current literature around methods for conducting systematic reviews, encompassing reviews of prevalence, prognosis, diagnosis and intervention effectiveness. In a field of methodology where some guidance is collated in well-known handbooks, but other areas are scattered among the methods literature, and where development of new methods is relatively fast-moving, this map of available guidance and methods studies is likely to be very useful to researchers new to systematic reviews or branching out into different review types. The authors are experienced experts in this field, and the review is well-written and

clear.

In my understanding, this paper aims to fulfil two roles: (1) a toolkit that enables authors of systematic reviews to navigate out to key methods resources, and (2) a brief summary of methods guidance for each review type. It may be helpful throughout the paper, including in the abstract and conclusions, to separate out these two roles more clearly. I have some suggestions to improve the transparency and usability of the paper from each of these perspectives.

Major recommendations

As a review and toolkit of methodological resources, I would make the following suggestions:

- **Methods:** The authors could provide more transparent detail on the selection process for included documents, including specifying which organisational websites were searched, how many resources were identified, and if any were excluded. Although this is intended as a mapping, rather than systematic review, some additional detail would help readers to understand how the guidance documents were selected, and why perhaps some resources they may have used before are not listed.
- Importantly, more detail could be presented on how the authors selected the most “relevant” or “best” resources, especially where there may have been multiple candidate documents. For example, were they most recent, most comprehensive, those endorsed by a credible organisation, most introductory, most rigorous methods? The role of the authors’ expert judgement in these decisions should be stated explicitly. Note that I am not suggesting that a different process should have been used, just suggesting a more detailed description of the judgement process.
- **Results:** As a navigation guide, I found that citing the resources within the text did not tell me everything I wanted to know. Also, the key resources are mixed in the reference list with exemplar reviews and other citations. A table outlining the recommended resources by category with hyperlinks to each resource would be helpful.
- As a reader, I would find it helpful for the authors to draw the distinction between the different types of resources cited. For example, some were synthesised, ‘best practice’ guidance intended for use by authors (such as the Cochrane or JBI Handbooks). Some were reporting guidelines (which often list good practice but are not intended to provide detailed guidance on the conduct of reviews). Some were primary methods studies (e.g. measuring the prevalence or impact of a method, which may describe or evaluate possible methods options but are not intended as guidance on the ‘best’ available methods for authors. All may be useful in different ways.

As a brief summary of methods guidance, I would make the following suggestions:

- **Methods:** It would be helpful to include a brief description of your role in selecting and writing this summary of guidance, for example, stating briefly that key areas of methods were summarised, and how the methods to be highlighted were selected, especially where multiple and potentially conflicting sources were available (e.g. major methods applicable to all reviews, key differences between review types, expert opinion about interesting or important advice).
- It might be helpful to note that the summary of guidance presented is just that, a brief summary, and that authors who are new to systematic reviews should consult the more

detailed documents for complete guidance.

- **Results:** It might be particularly helpful to note where there is disagreement in the literature in relation to particular areas of guidance. For example, comments that searching the grey literature is not useful for effectiveness reviews, or that risk of bias tools should be adapted for each review, which may be in disagreement with some of the major guidance handbooks and reporting guidelines. I'm not arguing about these specific items, just noting that some methods choices have been made by the authors, and it may be helpful to make this transparent.

Minor recommendations

- **Title and abstract:** It would be helpful to briefly define the scope of this review, especially in the context of a journal such as F1000Research, which publishes in a wide range of scientific fields. For example, using a term such as 'systematic reviews in health care', and noting that the review looks at prevalence, prognosis, diagnosis and health interventions (and not other areas relevant to health such as environmental exposure).
- **Competing interests:** The authors could note any roles with Cochrane and the GRADE Working Group. These are not necessarily financial conflicts, but may be relevant to the authors' decision to recommend guidance from these sources (which of course I agree with!).
- **Results:** In the first two paragraphs, you discuss some specific points in relation to resources available for diagnostic and prognosis reviews. As all review types appeared to cite both guidance handbooks and additional primary methods studies, I wasn't clear on the point you were trying to make in this section.
- There were some key guidance documents that I would have expected to see, although there may be good reasons not to include them, such as:
 - General guidance on systematic reviews from the US Institute of Medicine and the Centre for Reviews and Dissemination at the University of York (both are older documents, so this may be the reason).
 - Tools to assess the risk of bias in reviews, such as ROBIS or AMSTAR, which may enable reflection by authors on their choice of methods, in a similar manner to reporting guidelines.
 - Resources on the development of protocols (as distinct from study registration), such as Chapter 1 of the Cochrane Handbook.
 - Guidance on synthesis in the absence of meta-analysis, such as Chapter 12 in the Cochrane Handbook and papers by (Campbell, McKenzie *et al.* (2020)¹) and (Popay, Roberts *et al.* (2006)²). This may be relevant to synthesis, assessment of heterogeneity and GRADE.
 - Guidance on network meta-analysis, such as Chapter 11 of the Cochrane Handbook and multiple journal articles, as well GRADE guidance and CINeMA.
 - Two papers from the GRADE Working Group (published this week!) on the use of GRADE for diagnostic test accuracy studies published this week.
- In the text on risk of bias and Table 2, is it worth noting that the number of items in RoB 2 varies depending on the effect of interest and the included study designs?

- In Table 3, as there are more variations on the measures to combine (e.g. Ratio of Means), and statistical methods for meta-analysis of efficacy of interventions (including the inverse-variance method, which is mentioned for other types), could it be helpful to note in the table somewhere that these are common characteristics, not an exhaustive list?
- When discussing fixed effect and random effects models, it may be helpful to note that they differ in relation to assumptions and heterogeneity, as you mention something briefly about analysis using random-effects models later in the paper that relies on understanding this.
- On page 9, under “Efficacy of intervention reviews”, I think hazard ratios are listed in error against binary outcomes as well as time-to-event outcomes. Perhaps you meant risk ratios?
- In the section on assessing certainty of the evidence, it would be helpful to describe the GRADE approach in the first paragraph (it is currently named without description in the third paragraph).
- **Discussion:** I would also acknowledge under limitations of this review that the selection of resources and summary of guidance were informed by expert opinion, and that others may have selected different resources or made different recommendations.
- **References:** 1: I think there may be an error – should this be a reference to the 2019 edition of the Cochrane Handbook, rather than an older paper by Higgins in *Cochrane Methods* on RoB 2? I assume this based on its use in the Introduction as a general reference about review methods, and also on the availability of general Handbooks in para 2 of the Results.
- 20: could be based on the references to individual chapters, i.e. Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*.
- 63: It's great to see the Spanish Manual GRADE cited. If I am correct, this is a translation of the 2013 GRADE Handbook (apologies if I am wrong). Would it be helpful to cite some of the more recent GRADE papers as well, as they reflect the most up to date guidance? I note you also cite the paper on using standard language to express GRADE, as well as the recent Cochrane Handbook Chapter, so perhaps this is sufficient.

References

1. Campbell M, McKenzie J, Sowden A, Katikireddi S, et al.: Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ*. 2020. [Publisher Full Text](#)
2. Popay J, Roberts H, et al.: Guidance on the conduct of narrative synthesis in systematic reviews. *ESRC methods programme*. 2006. 47-71

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

No

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: I am an Associate Editor of the Cochrane Handbook for Systematic Reviews, and an Editor with Cochrane Public Health.

Reviewer Expertise: In presenting my comments, I acknowledge that my own experience focuses on methods around systematic reviews of the effects of interventions, and so I cannot comment with expertise on the details of methods described for other review types.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 19 Jul 2020

Marta Roqué, Iberoamerican Cochrane Centre - Sant Pau Biomedical Research Institute (IIB-Sant Pau), Barcelona, Spain

- Methods: The authors could provide more transparent detail on the selection process for included documents, including specifying which organisational websites were searched, how many resources were identified, and if any were excluded.

ANSWER: Thanks for this suggestion. The text lists the main organisational websites we checked to identify guidelines ("Cochrane, Joanna Briggs Institute, European Network for Health Technology Assessment (EUNETHTA), Enhancing the Quality and Transparency of Health Research (EQUATOR) network, Grading of Recommendations Assessment, Development and Evaluation (GRADE)", as suggested by the peer reviewer. Also, the text is now more explicit about the process followed to select resources. See answers to comments below regarding this specific issue. More detailed description of the resource selection process has been added to the text (see detailed modifications in the next comments)

- Importantly, more detail could be presented on how the authors selected the most "relevant" or "best" resources, especially where there may have been multiple candidate documents.

ANSWER: Thanks for this suggestion. A paragraph was added to the eligibility criteria section "The resources were selected based on the authors expert judgement, prioritising those resources which were endorsed or part of a guideline from the organisations cited

above, and those which were more recent. "

- Results: As a navigation guide, I found that citing the resources within the text did not tell me everything I wanted to know.

ANSWER: We have added a table in the Appendix where the references and resources considered in the review are classified into guidelines to conduct SR (or chapters of those guidelines), reporting guidelines, primary methods papers and references to examples used in the manuscript.

- Methods: It would be helpful to include a brief description of your role in selecting and writing this summary of guidance,

ANSWER: Thanks, we will be more explicit on this issue. A sentence has been added to the data selection and extraction section "For each pre-defined section, the authors selected and summarized the methods that were considered to be more rigorous and widely accepted, prioritizing major methods applicable to all reviews over more controversial methods, or methods which required highly specialized knowledge. The text organises the results pedagogically with the aim to highlight key differences between review types, present the key characteristics of each method, and be a comprehensive tool that contains the most relevant advice based on the authors judgement."

- It might be helpful to note that the summary of guidance presented is just that, a brief summary, and that authors who are new to systematic reviews should consult the more detailed documents for complete guidance.

ANSWER: Thanks for the suggestion, we have clarified this point in the text. A sentence has been added to the key results section of the discussion, stating "This project does not aim to be a standalone tool for a researcher to find complete guidance on how to conduct and report a review, but rather it aims to be a signpost pointing out to the resources where researchers may find in depth guidance to develop their reviews."

- Results: It might be particularly helpful to note where there is disagreement in the literature in relation to particular areas of guidance.

ANSWER: As far as possible, we have avoided presenting controversial advice, although we are aware that any topic can be approached differently by different researchers or even institutions. It could be quite daunting to be comprehensive in identifying all the controversies in the different steps of conducting a review, as the volume of publications on methods for reviews is extremely large and diverse. Additionally, we think that the discussion or even the identification of issues where controversy exists may fall outside of the project scope, as it might reduce the usefulness to a new researcher which needs to find clear guidance on a topic, even if there are other alternative methods available. We have stressed the subjectiveness and risk of implicit selection biases in the discussion "An inherent limitation of this project is its methodology based on a selection of resources and summary of guidance informed by expert opinion, which may be susceptible to implicit selection biases or lack of comprehensiveness. "

- Title and abstract: It would be helpful to briefly define the scope of this review, especially in the context of a journal such as F1000Research, which publishes in a wide range of scientific fields.

ANSWER: Thanks, this is a very useful suggestion. Title has been modified to "Toolkit of methodological resources to conduct systematic reviews in health care: reviews on prevalence, prognosis, diagnosis and interventions". The concept has also been introduced in the abstract ("This work identifies and describes the most relevant methodological resources to conduct high-quality reviews that answer health care questions regarding

prevalence, prognosis, diagnostic accuracy and effects of interventions")

- Competing interests: The authors could note any roles with Cochrane and the GRADE Working Group.

ANSWER: A sentence was added to the Data selection and extraction "The authors are members of CIBERESP (Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública - Biomedical Research Center Network of Epidemiology and Public Health), hold active roles within Cochrane and the GRADE Working Group,"

- Results: In the first two paragraphs, you discuss some specific points in relation to resources available for diagnostic and prognosis reviews. As all review types appeared to cite both guidance handbooks and additional primary methods studies, I wasn't clear on the point you were trying to make in this section.

ANSWER: We realize we didn't make our point clear, that was to illustrate what guideline handbooks had been identified by type of review. We've rewritten both paragraphs aiming to be more clear. "We identified guidance handbooks, primary studies and reporting guidelines as a result of the bibliographic searches. The resources selected are presented in the Appendix.

We have identified methodological guidelines dedicated to the development of prevalence SRs¹⁴, global prognosis¹⁵, and prognostic factor SRs^{16, 17, 18}.

During the performed search, we identified methodological manuals to develop prognostic model SRs in the series of publications from the PROGRESS project¹⁹, and in the resource compilation from Cochrane's Prognosis Methods Group."

- There were some key guidance documents that I would have expected to see,

ANSWER: .Thanks for the bringing forth this issue, which was overlooked in the manuscript. The issue of narrative synthesis of results has been cited in several sections the paper (statistical synthesis, quality of evidence, results report). An explicit mention to narrative synthesis has been added to the Statistical synthesis section "In those cases when a quantitative synthesis is precluded, the SR will be restricted to a narrative synthesis. A narrative synthesis should not simply summarize the findings from the included studies in order to draw conclusions about the body of evidence, but instead should be a more formal process which includes a formulation of the theory of how the intervention works, why and for whom, the exploration of the relationships in the data, and the assessment of the robustness of the synthesis. 59". Certainty of evidence from narrative synthesis has been mentioned "Certainty of evidence can be assessed too when no quantitative synthesis is possible. 65". Also, the SWiM paper has been cited in the REsults report section "Additionally, the SWiM guideline is available for reporting intervention SRs where the effects of interventions are synthetised narratively without metanalysis, focusing on the key features of narrative information synthesis (grouping of studies, presentation of data and summary text, and appropriate discussion of limitations of this type of synthesis)."

Network metanalysis have been mentioned in the Statistical methods and Quality of evidence sections. A cite to Chapter 11 in the handbook has been added to both sections Thanks, we've added the GRADE diagnostic references to the paper. These references were incorporated to the corresponding Quality of evidence section, modifying the paragraphs on diagnostic SR.

- In the text on risk of bias and Table 2, is it worth noting that the number of items in RoB 2 varies depending on the effect of interest and the included study designs?

ANSWER: Thanks, we've added this information. A sentence has been added to the risk of bias section ("The number of questions may vary, depending on the effect of interest and the design of the study assessed"). A footnote has been added to Table 2 "The number of items in the risk of bias tools may vary depending on the effect of interest and the included study designs, as well as the addition or suppression of index questions by the researchers to tailor the tool to the SR."

- In Table 3, as there are more variations on the measures to combine (e.g. Ratio of Means), and statistical methods for meta-analysis of efficacy of interventions (including the inverse-variance method, which is mentioned for other types), could it be helpful to note in the table somewhere that these are common characteristics, not an exhaustive list?

ANSWER: Thanks, we've corrected the oversight of not mentioning the inverse variance method for continuous outcomes in the efficacy of interventions SR, and have clarified in the text that the table is by no means exhaustive. The text introducing table 3 now reads "Table 3 shows a non-exhaustive compilation of the main characteristics of the meta-analysis methods and the main software commands for each type of SR". The inverse-variance method has been added to the efficacy of interventions row

- When discussing fixed effect and random effects models, it may be helpful to note that they differ in relation to assumptions and heterogeneity, as you mention something briefly about analysis using random-effects models later in the paper that relies on understanding this.

ANSWER: This is certainly an important point. Although there is already an explicit link between choice of model and heterogeneity in the previous paragraph ("the evaluation of the existing clinical and statistical heterogeneity in the set of studies, which will inform us /.../ 2) what meta-analysis model we should apply"), we agree to further stress this point as suggested. We've modified the existing sentence on random and fixed-effects models "Both models are based on different assumptions regarding distribution of effects and heterogeneity across studies, and they differ in their application and interpretation"

- On page 9, under "Efficacy of intervention reviews", I think hazard ratios are listed in error against binary outcomes as well as time-to-event outcomes. Perhaps you meant risk ratios?

ANSWER: You're right, it should have read 'risk ratios' instead of 'hazard ratios'. hazard ratio' substituted by 'risk ratio'

- In the section on assessing certainty of the evidence, it would be helpful to describe the GRADE approach in the first paragraph (it is currently named without description in the third paragraph).

ANSWER: Thanks, we agree. We've added an explicit reference to the GRADE system "Certainty of evidence is best evaluated with the GRADE system. Certainty in the obtained estimates for each one of the key SR outcomes or factors is classified as high, moderate, low or very low"

- Discussion: I would also acknowledge under limitations of this review that the selection of resources and summary of guidance were informed by expert opinion, and that others may have selected different resources or made different recommendations.

ANSWER: While we have strived to draw a unbiased selection of the best resources, it is right to point out this potential limitation. We've added the sentence "An inherent limitation of this project is its methodology based on a selection of resources and summary of

guidance informed by expert opinion, which may be susceptible to implicit biases or lack of comprehensiveness"

- Reference 1: I think there may be an error – should this be a reference to the 2019 edition of the Cochrane Handbook, rather than an older paper by Higgins in Cochrane Methods on RoB 2?

ANSWER: Yes, you are right. References 1 and 56 were interchanged, and have now been corrected. New reference 1 and new reference 56

- Reference 20: could be based on the references to individual chapters, i.e. Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.

ANSWER: Thank you, the reference has been modified as suggested

- Reference 63: It's great to see the Spanish Manual GRADE cited. If I am correct, this is a translation of the 2013 GRADE Handbook (apologies if I am wrong). Would it be helpful to cite some of the more recent GRADE papers as well, as they reflect the most up to date guidance? I note you also cite the paper on using standard language to express GRADE, as well as the recent Cochrane Handbook Chapter, so perhaps this is sufficient.

ANSWER: We wish to keep the original 2013 reference, as the most comprehensive manual on GRADE. However, we will substitute the reference of the Spanish version for the English version. Reference 63 substituted for the reference to the original version

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research