A practical guide:

Screening for Systematic Reviews

covidence

Introduction

Welcome to this practical guide to screening for systematic reviews.

Whether you're an experienced researcher or just starting out, this guide can help you plan and manage the screening processes for your systematic review.

Inspired by the knowledge of hundreds of researchers, this guide compiles best practices and tips from the global systematic review community. It features clear definitions and practical advice.

We hope this guide becomes an essential part of your research journey.

About the author

We are Covidence. Launched in 2014, Covidence is a not-for-profit world leading Software as a Service (SaaS) platform. Our platform enables health and science research teams to rapidly synthesise and uncover actionable insights from the mountains of research produced around the world. Leading institutions worldwide use Covidence to create the knowledge that shapes our society.

If you find this guide helpful, please share it with your community so everyone can benefit. Feel free to use the pictures and drawings in your own content. We'd appreciate it if you could include a shout-out: 'Diagrams and illustrations courtesy of Covidence,' along with a hyperlink to the eBook whenever you can. Thanks for spreading the word!

Contents

01

Importance of screening

Introduction

Why is screening important?

Screening is an important step in any type or size of literature review. The process of screening ensures that the studies in the review are relevant to the review question. It also ensures that the review is based on the highest-quality evidence and adheres to pre-specified criteria. Screening can be extremely resource intensive.

Screening is critical to ensure:

- **Relevancy:** Only studies addressing the research question are included in the review. Irrelevant studies are removed or excluded.
- **Consistency:** Screening uses pre-specified eligibility (inclusion and exclusion) criteria to reduce bias and provide transparency. These criteria should be pre-specified in your review protocol.
- **Replicability:** Pre-specified screening procedures (number of reviewers, blinding, conflict resolution) allow other researchers to replicate the review.
- **Quality:** Screening can be used to eliminate low-quality or poorly designed studies, which may introduce bias and lead to misinterpretation of review findings.

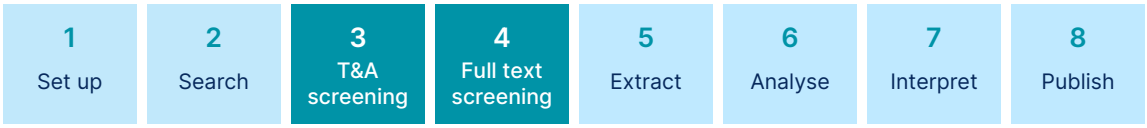
Where does screening fall in the systematic review process?

Screening usually takes place between the literature search of sources/databases and data extraction. There are usually two steps to screening:

- Title and abstract screening - Identifies and removes irrelevant studies
- Full text screening - Identifies studies for data extraction through clarification of eligibility criteria by reading full text of study

Producing an output that supports analysis and comparison of extracted data requires the completion of:

- standardised data extraction for each study
- quality assessment (risk of bias) for each study



Review workflow

By adhering to a rigorous screening process, systematic reviews maintain their integrity, providing robust and trustworthy evidence for decision-making.

02

Screening and the review protocol

Screening and the review protocol

What are the considerations for screening in the review protocol?

The protocol should describe the approach that will be used to identify potentially relevant studies (title/abstract screening) and to select included studies (full text screening). You ensure the transparency of your review by providing these details. Transparency is essential to reduce bias within the review process.

The eligibility (inclusion and exclusion) criteria to be used for screening decisions must be documented in the protocol. It is mandatory to report these criteria in PRISMA reporting guidelines, the AMSTAR checklist and the Cochrane Handbook (Chapter 3).

> “Decisions about which studies to include in a review are among the most influential decisions that are made in the review process and they involve judgement.”
>
> - Cochrane Handbook 4.6.4

The review protocol should include details on:

- The number of reviewers (including names, if available) required to screen each reference for eligibility at title and abstract and full text screening stages.
- Which tools, if any, will be used to support screening (e.g. Covidence).
- Whether or not screening will be conducted independently (blinded).
- How, and by whom, conflicts will be resolved at title and abstract and full text screening stages.
- Whether, or not, inter-rater reliability will be assessed.

Why is blinding important during screening?

Independent screening is recommended for Cochrane and other systematic literature reviews.

Without blinding, the personal preferences and biases of the reviewer may influence their selection decisions which increases the risk of selection bias. Blinding of reviewers adds to the rigour and therefore the credibility of the review process.

Your protocol should detail the number of reviewers required to screen each reference and whether or not these reviewers will work independently. A justification should be provided if independent screening is not used.

Independent screening (blinding) is not essential for all review types but is crucial for reducing bias and ensuring reliability and objectivity of the review process.

Example protocol

“Selection process 11b State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)” PRISMA (P) statement

MM and GG will independently screen titles and abstracts and full text for relevancy. Where disagreements can not be resolved, HH will act as an arbitrator and make the final screening decision.

Deviating from the review protocol during screening

It is fairly common to make changes that relate to screening in the systematic review protocol. There may be changes to the review team allocated to screening, or changes to the proposed methods for screening (blinding versus unblinded; single- versus dual-reviewer screening).

Remember to clearly document any agreed changes to the review protocol with a clear rationale for the change. This is key for transparency and accountability. If the protocol has been registered with a protocol registry, such as PROSPERO, update it there too.

PRISMA for systematic review protocols (PRISMA P) - Item 4

If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments

03

Who should be involved in screening?

Who should screen?

Who should ideally be on the screening team?

A well-rounded team is crucial for conducting a systematic review. A systematic review team can be of any size but ideally should be more than a single reviewer. If you are a student then the review team might only consist of you and your supervisor. If you are in a more formal review team you might have a mix of content experts, methodologists, as well as novice/trainee reviewers. Here are some examples of who could be involved in the screening team for your review:

- **Lead reviewer:** Usually oversees the review process and provides content expertise.
- **Content experts:** These team members provide an in-depth understanding of the subject matter, intervention, context or population of interest.
- **Methodologist:** Individuals with a sound knowledge of systematic review processes and frameworks.
- **Reviewers (Screeners):** Team members who screen the titles and abstracts and full text studies against the review eligibility criteria. They may, or may not, have content expertise.

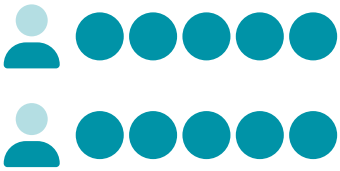
If you are a student then you may be the lead reviewer and your supervisor may be a co-reviewer who is monitoring your work. Your supervisor may be the content expert.

Remember to document the team and their roles in the review protocol.

Why is the number of reviewers important?

The review team will need to decide whether to have one or two reviewers screen each study. This is often known as single- or dual-reviewer screening. Having more than one reviewer undertake screening will increase the reliability and credibility of the screening process. It avoids references being accidentally missed and reduces the risk of introducing personal biases and subjectivity. Any conflicts can be resolved by a third reviewer or the review team.

Single reviewer



Dual-reviewer

Cochrane recommendation

"Use (at least) two people working independently to determine whether each study meets the eligibility criteria." Cochrane Handbook (Chapter 4.6.4)

Although dual-reviewer screening increases the methodological rigour of the review, it is resource intensive.

Having reviewers screen independently, or blinded to the other reviewer's vote also maintains the rigour of the review. It prevents potential personal biases and preferences from influencing another reviewer.

Who should ideally be on the screening team?

There are some screening alternatives that can be used, some of which can save time but they do reduce the rigour of the review and will increase the risk of introducing biases.

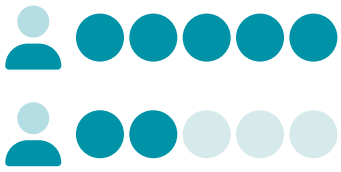
- **Proportional screening:** A subset of studies is independently reviewed by a second (or multiple) reviewer(s) to validate the decisions of the primary reviewer. This approach is used to optimise resources whilst maintaining accuracy and consistency in the study selection process. It is a useful method with large reviews or where there is a time constraint.

- **Consensus screening:** Reviewers screen together, discussing each decision in real-time. This is feasible for smaller reviews but will introduce bias as blinding has been eliminated.

- **Single reviewer with verification:** A single reviewer screens studies, and a second reviewer checks a subset for accuracy. The second reviewer is not blinded and the omission of blinding will introduce bias.This method should only be used in time-constrained situations.

How does proportional screening work?

Proportional screening is a useful alternative to dual-reviewer screening as it saves time whilst maintaining the rigour of blinding reviewers. The disadvantage is that a single reviewer might miss relevant studies when they screen alone. If the review team uses this method, then the protocol should clearly state the percentage of studies that will be screened in dual-reviewer mode and if, and when, inter-rater reliability will be assessed.



Proportional screening

How to conduct proportional screening:

How to conduct proportional screening:

1. **Primary reviewer:** Screens all titles, abstracts, and/or full texts for inclusion or exclusion based on the review eligibility criteria.

2. **Second reviewer(s):** The second reviewer/s independently screens a proportion (e.g. 10-20%) of the total studies. The proportion is typically determined by the project's size, available resources, and the need for quality assurance.

3. **Conflict resolution:** Conflicts are resolved through discussion or by involving a third reviewer.

4. **Checking inter-rater reliability (IRR):** It is important to check reliability before moving to single-reviewer screening. If the IRR (agreement between reviewers) is high, the primary reviewer's screening can continue without further validation. If agreement is low, proportional screening may need to be continued, or a dual-reviewer approach used for all remaining studies.

Cochrane recommendation

Cochrane Rapid Review Guidance

Title and Abstract Screening

- Using a standardised title and abstract form, conduct a pilot exercise using the same 30-50 abstracts for the entire screening team to calibrate and test the review form.
- Use two reviewers for dual screen of at least 20% (ideally more) of abstracts, with conflict resolution.
- Use one reviewer to screen the remaining abstracts.
- Use a second reviewer to screen all excluded abstracts, and resolve conflicts.

Full Text Screening

- Using a standardised full text form, conduct a pilot exercise using the same 5-10 full-text articles for the entire screening team to calibrate and test the review form.
- Use one reviewer to screen all included full text articles.
- Use a second reviewer to screen all excluded full text articles.

Tips

- Establish your research team early in the review process.
- Assign tasks to reviewers including roles and responsibilities for screening.

04

Preparing references for screening

Who should screen?

Make sure that the references are as complete as possible (title and abstract are included, if available) before you start screening. This is also the time to reconsider your search strategy if you think that you have too many or too few references. It's good practice to identify the number of references that have been retrieved from each source (database, website etc.). Make sure that you have removed duplicates (to avoid time wastage in screening references more than once).

Make sure references are complete

When you are exporting references from your source databases or reference manager make sure that you have selected 'all fields'.

At a minimum you should export the:

- Reference (citation)
- Abstract
- Digital Object Identifier (DOI) link

What to do if you think you've got too many or too few studies

Your search may be too broad. This means that you may have many thousands of irrelevant studies. Your search may be too narrow which means that you may be missing relevant studies. There may be an error in the terms or limitations used in the search strategy.

In these circumstances, the team should revisit the review protocol and check the search strategy. You may need to revise the strategy or add/remove some limitations. You must document and justify any changes that you make to your protocol.

Identifying sources for PRISMA

The methodology and results of systematic reviews need to be reported with enough information to allow users to assess the trustworthiness and applicability of the review findings. The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement facilitates transparent and complete reporting of systematic reviews and supports adherence to standardised reporting guidelines in systematic reviews (Page, 2021).

It is good practice to record the number of references retrieved from each source. A source can be a database (e.g. PubMed, ERIC), a website, government documents, or grey literature. This information can be added to the PRISMA flow diagram. This level of detail improves the transparency and reproducibility of the review.

What is a duplicate?

Different databases can index the same studies which results in multiple instances of the same reference appearing in the search results when you import references from multiple databases. These are known as **duplicates.** Identifying duplicates can sometimes be difficult due to minor variations in metadata and variability in presentation between databases.

Example of duplicate references

Martha Meta, Grace Graph, Peter Piper. Vitamin C for prevention and treatment of the common cold in older adults. New England Journal of Medical Sciences. 2025; 234 (1):231-245.

Meta M, Graph G, Piper P. Vitamin C for prevention and treatment of the common cold in older adults. NEJMS.2025; 234(1):231-45. doi:10.123456789xx.

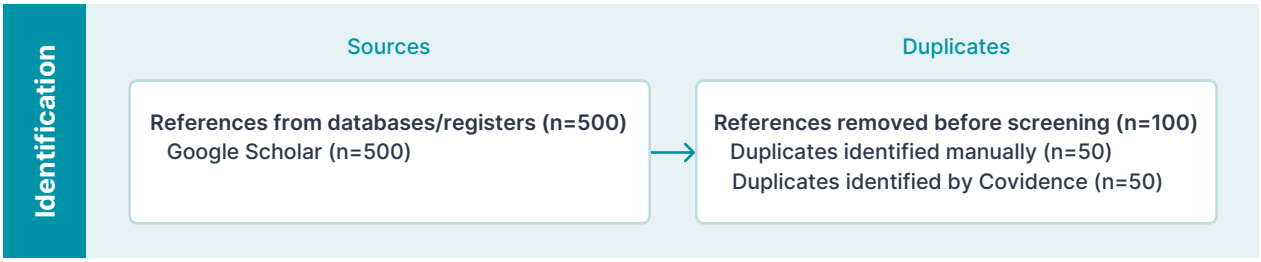
Duplicates can be identified by checking key features of the references, such as:

- **Titles:** Look for titles that are identical or very similar.
- **Authors:** The same set of authors listed in the same order.
- **Publication year:** Matching years of publication.
- **Journal name:** Check for the same journal name and volume, issue and page numbers.
- **Digital Object Identifier (DOI):** The same DOI is a clear indicator of a duplicate.

What is deduplication?

Deduplication is the identification and removal of duplicate records from the pool of references retrieved during the literature search.

Deduplication reduces the workload during screening. It can be done prior to screening using automation tools in systematic review software such as Covidence or in reference managers such as Endnote/Zotero. Removal of duplicates can also be done during the screening process (manually identified duplicates). The number of duplicates identified by automated and manual methods should be documented in the PRISMA figure.



Tips

- Make sure that the exports from databases are complete or, at a minimum, contain the reference (citation), abstract and Digital Object Identifier (DOI) link.
- Always check duplicates removed by automation tools for false positives.
- Be careful not to confuse a duplicate with multiple publications from the same study.

Example of multiple publications from the same study

Martha Meta, Grace Graph, Peter Piper. Vitamin C for prevention and treatment of the common cold in older adults. New England Journal of Medical Sciences. 2025; 234 (1):231-245 (Page, 2021).

Martha Meta, Grace Graph, Peter Piper. Vitamin C for prevention and treatment of the common cold in older adults. A 6-months follow up study. New England Journal of Medical Sciences. 2025; 234 (7): 127-131 (Page, 2021).

05

Title and abstract screening

Title and abstract screening

Title and abstract screening is the first crucial step to identify potentially relevant studies. It is the process of sifting through the potentially overwhelming volume of studies identified in the literature searches, to remove irrelevant references and to provide a list of references that appear to meet the review eligibility criteria. This list will then need to be screened by full text review.

It could be very costly and time consuming if the full text for every potential study had to be retrieved for title and abstract screening.

Allocation of roles

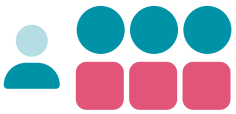
The review team must decide who is going to screen the title and abstract studies. Everyone in the team may be involved or only selected team members. Title and abstract screening in a systematic review should ideally be undertaken by two independent reviewers. The number of reviewers (and names if possible) required to screen each reference should be documented in your protocol.

Allocation of references

Some review teams like to allocate specific references to each reviewer (e.g Authors A-C; ID number 1-50). Other teams might ask reviewers to screen a specific number of references (Any 50 references).

One-step vs two-step screening

Studies can be screened using one- or two-step methods. One-step screening reviews titles and abstracts simultaneously. The two-step process screens titles first and then screens the remaining studies using both titles and abstracts. Either way is fine as long as the process is documented in your protocol.



One-step screening

Two-step screening

Example one-step screening

Step 1 - Screen titles and abstracts simultaneously

> *Reference*: Vitamin C for preventing and treating the common cold (2023) Miles MM & Green GG. J Adv Systematic Revs. 10(6):12-21.
>
> *Abstract*: Objective: To assess the efficacy of vitamin C in reducing the incidence and duration of the common cold.

Methods: We conducted a randomised controlled trial involving 500 adults who were assigned to receive either 1000 mg of vitamin C daily or a placebo for 12 weeks. Participants were monitored for the occurrence of colds at 1 week, 1 month, and 3 months. Additionally, the duration of colds in days was recorded for those who contracted the illness.

Results: At 1 week, 1 month, and 3 months, there was no significant difference in the number of colds between the vitamin C and placebo groups. However, the duration of cold symptoms was significantly shorter in the vitamin C group (mean difference of 1.2 days, p=0.03).

Conclusion: Daily vitamin C supplementation does not reduce the incidence of the common cold, but it may shorten the duration of symptoms in those who become ill. Further research is needed to confirm these findings.

Example two-step screening

Step 1 - Initial screen by reference/title only

> *Reference*: Vitamin C for preventing and treating the common cold (2023) Miles MM & Green GG. J Adv Systematic Revs. 10(6):12-21.

Step 2 - Go though remaining titles and abstracts

> *Reference*: Vitamin C for preventing and treating the common cold (2023) Miles MM & Green GG. J Adv Systematic Revs. 10(6):12-21.
>
> *Abstract*: Objective: To assess the efficacy of vitamin C in reducing the incidence and duration of the common cold.

Methods: We conducted a randomised controlled trial involving 500 adults who were assigned to receive either 1000 mg of vitamin C daily or a placebo for 12 weeks. Participants were monitored for the occurrence of colds at 1 week, 1 month, and 3 months. Additionally, the duration of colds in days was recorded for those who contracted the illness.

Results: At 1 week, 1 month, and 3 months, there was no significant difference in the number of colds between the vitamin C and placebo groups. However, the duration of cold symptoms was significantly shorter in the vitamin C group (mean difference of 1.2 days, p=0.03).

Conclusion: Daily vitamin C supplementation does not reduce the incidence of the common cold, but it may shorten the duration of symptoms in those who become ill. Further research is needed to confirm these findings.

Using software to tag studies

Some systematic review management software allows you to tag studies with key words or phrases that can then be filtered. These might include tags for studies that need translation or you might use tags to allocate studies to a team for screening (Team A, Team B). Some software such as Covidence allows you to tag studies that are possibly randomised controlled trials. These tools can speed up screening.

Identifying fraudulent studies/retractions/errata

It is important to be aware of studies that have been retracted and of published errata. Studies are often retracted for errors, fraudulent reporting, plagiarism or ethical violations and including these studies can compromise review integrity and credibility. Their inclusion may skew findings and lead to erroneous conclusions and incorrect recommendations.

Some reference managers such as Zotero check for retractions. You can also search databases such as Retraction Watch, PubMed's retraction notices, and monitor journal websites.

Identifying study errata is important when screening. You should not exclude these citations. Corrected data may affect your review findings and it is important that your review is based on the most accurate and recent information. You should document if published errata have resulted in the amendment of data in your review.

Voting in title and abstract screening

During title and abstract screening the most common options for voting are:

- **Yes** - moves studies on to full text review.
- **No** - these studies are irrelevant and require no further action.
- **Unsure/Unclear/Maybe** - moves studies to full text review. There might be missing information or lack of clarity as to whether the eligibility criteria have been met.

A conflict is generated where there are disagreements in voting Yes vs No or Maybe/ Unsure/Unclear vs No. We discuss conflicts and how to resolve them in more detail here.

All voting should be documented in a PRISMA flow diagram (or similar).

Is it necessary to give a reason for exclusion during title and abstract screening?

It is generally sufficient to provide details of the numbers of studies that are excluded at title and abstract screening. Some teams prefer to identify reasons for exclusion for title and abstract screening as well as full text screening. Whichever method is used should be documented in the review protocol.

Piloting screening

For most reviews it will be worthwhile to pilot test the eligibility criteria on a sample of studies (e.g six to eight articles, including ones that are thought to be definitely eligible, definitely not eligible and maybe eligible). The pilot test can be used to refine and clarify the eligibility criteria, train the screening team and ensure that the criteria can be applied consistently by more than one person. This will reduce the number of conflicts and speed up screening.

Tips

- Have some brief instructions for the reviewers undertaking screening. This keeps the team aligned, especially if there are complicated definitions or conditions.
- Document the processes for screening in your review protocol.
- Pilot the screening to make sure that everyone on the team understands the processes and eligibility criteria.

06

Full text screening

Full text screening

So why is full text screening important?

Full text screening follows title and abstract screening where the irrelevant studies were removed. It identifies the studies requiring data extraction. Full text screening is an important step in the review process because this screening stage:

- **Ensures relevancy**: Full text screening checks whether the article meets the eligibility (inclusion and exclusion) criteria of the review. Sometimes the information provided in the title and abstract is insufficient to determine relevance or may contain inaccurate information. Full text screening enables a comprehensive evaluation.
- **Identifies potential biases**: Some studies could report additional data or outcomes in the main body of the paper that were not mentioned in the abstract. Alternatively, some outcomes that were listed in the abstract may not be reported in the full text.
- **Improves quality assessment**: Full text screening ensures that methodological as well as outcome data are screened. This is important if methodological quality is a component of the eligibility criteria.

Full text retrieval

Once you start full text screening you need to be able to access and view the full text article or PDF. You cannot fully assess eligibility without reading the article or document. Some articles are free-to-access and others will need to be located via your library services or through inter-library loans. Some articles may require translation. If full text articles are not available then this should be documented in your report.

Many review teams will manage their PDF retrieval through reference management software. There are some useful browser extensions/Add-ins that can make retrieval easier especially when you link your University library services e.g Endnote Click, Unpaywall, and LibKey Nomad.

Allocation of roles

The team will need to decide who is going to screen the full text articles. Everyone in the team might be involved or only selected team members. Full text screening in a systematic review should ideally be undertaken by two independent reviewers. The number of reviewers (and names if possible) required to screen each reference should be documented in the protocol.

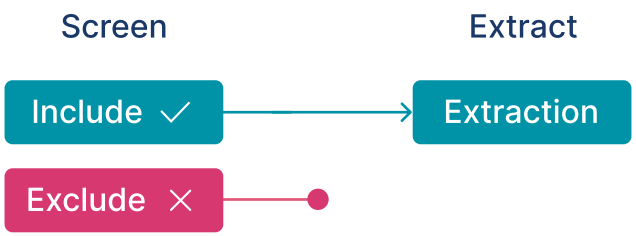
Allocation of studies

Some review teams allocate specific references to each reviewer (e.g Authors A-C; ID number 1-50). Other teams might ask reviewers to screen a specific number of references (Any 50 references).

Full text screening

During full text screening the reviewers are checking the article against the eligibility criteria. It is important that these criteria are readily available to the reviewers who are screening.

During full text screening the voting options are usually:



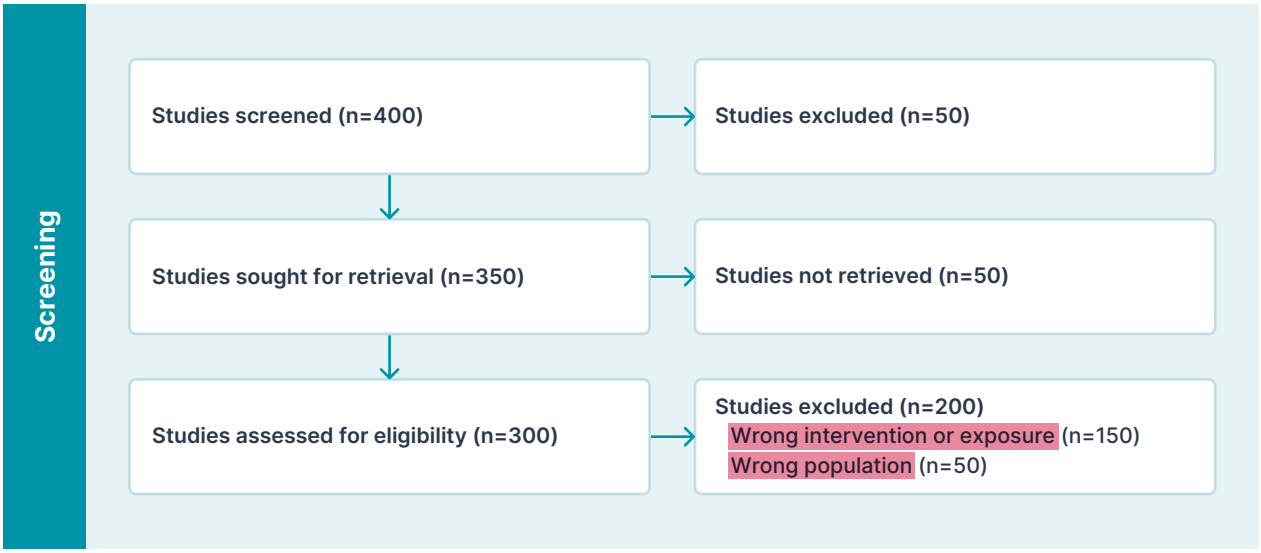
The same processes for resolving conflicts applies to full text screening as was described for title and abstract screening. Where there are conflicts (disagreements) in voting selection the options include a resolution between reviewers or the use of a third independent adjudicator where conflicts cannot be resolved.

Selecting reasons for exclusion

At least one explicit reason should be provided when excluding studies during full text screening. The lists of included and excluded studies must be based on studies rather than individual publications. Documenting excluded studies and providing the rationale for exclusion shows the reader that consideration has been given to these studies. The reason for exclusion and the number of associated studies are documented in the PRISMA figure.

Some common reasons for excluding studies are:

- Wrong intervention or exposure
- Wrong population
- Wrong study design
- Wrong dose
- Wrong setting



Resolving conflicts

The same processes for resolving conflicts applies to full text screening as has been described for title and abstract screening. Where there are conflicts (disagreements) in voting selection the options include a resolution between reviewers or the use of a third independent arbitrator where conflicts cannot be resolved.

Studies can have more than one valid reason for exclusion. To avoid conflicts based on reason for exclusion the team could order the suggested reasons into a hierarchy. When the team is screening the reason for exclusion should be selected in the same order and this will certainly minimise the number of conflicts generated.

Does title and abstract screening need to be completed to begin full text screening?

Some teams like to complete title and abstract screening before progressing on to full text screening.

Some teams take a ‘waterfall’ approach where full text screening can be started whilst title and abstract screening is still underway. This can significantly reduce the time to complete the review and can make use of the different skills within the team. It prevents delays when some team members were unable to complete earlier screening tasks on time.

Tips

- Have the eligibility criteria easily accessible to the reviewers who are screening.
- Retrieve the associated full text PDF or document before you start screening a study.
- Set up a hierarchy for the reasons for exclusion to minimise conflicts.

07

Resolving conflicts

Resolving conflicts

Conflicts arise in both title and abstract and full text screening during dual-reviewer screening. This is when there is a 'disagreement' between the two reviewers who are screening. The cause of the conflict is often a misunderstanding, simple oversight or a voting mistake on the part of one of the reviewers.

These conflicts can usually be resolved by a quick discussion. Arbitration by a third reviewer may be needed where the reviewers cannot agree.

Conflicts in title and abstract screening

Most systematic reviews will have three screening options during title and abstract screening:

- Yes
- Maybe/unsure/unclear
- No

Conflicts arise when one reviewer votes 'Yes' and the second reviewer votes 'No'; or when one reviewer votes 'Maybe' and the second reviewer votes 'No'.

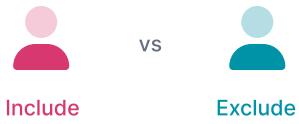
These conflicts tend to occur when one of the reviewers missed or misunderstood an eligibility criteria (inclusion/exclusion criteria), or accidentally selected the wrong option.

Refer back to the review protocol and eligibility criteria when resolving these conflicts. These conflicts are simply and easily resolved, usually by the reviewers who voted. Resolution can usually be done in a discussion by email, in-person or by video discussion. Some teams prefer a senior member of the team to resolve conflicts. If the conflicts can not be resolved by the two original voters then an independent third reviewer may be involved. How conflicts will be resolved and who will resolve them must be documented in your protocol.

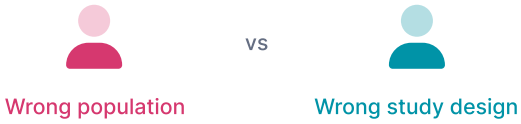
Conflicts in full text screening

In full text screening, reviewers are usually given the options to include or exclude a reference. A single reason for exclusion is required if a reference is excluded. This reason is included in the PRISMA flow diagram.

Conflicts can arise when reviewers disagree on whether to include or exclude a study.



Conflicts can also arise when both reviewers vote to exclude a reference but they provide a different reason for exclusion (PRISMA allows one reason per reference).



As with title and abstract screening, resolving conflicts is usually done by the two reviewers who voted and can be done via a quick email conversation or online call. Again these conflicts are usually due to reviewers having missed or misunderstood an eligibility criteria, or accidentally selecting the wrong option. Some teams prefer a senior team member to resolve conflicts at the full text stage as they may have more content/ methodological experience. Bringing in a third reviewer is a solution on the rare occasion when a consensus agreement cannot be reached by the two reviewers.

Refer back to the review protocol and eligibility criteria when resolving these conflicts.

Resolving conflicts can add time delays to your review so it is always a good idea to pilot both title and abstract and full text screening and make sure that the team understands the eligibility criteria and have a hierarchy of reasons for exclusion. In order to keep things moving, the team may decide to have a dedicated reviewer or reviewers to deal with conflicts which can be resolved at any time.

What if there is insufficient information to resolve a conflict?

Sometimes conflicts can not be immediately resolved because there is not enough information. Review teams may decide to identify these studies as 'Awaiting classification' until such time as the additional information is obtained from the study authors/team.

Tips

Remember, some studies have more than one valid reason for exclusion which accounts for the second type of conflict that is often observed in full text screening. A good solution here is to order your list of reasons for exclusion into a hierarchy. This means that when the team is screening and excluding studies they will always select the reason in the same order and this will certainly minimise the number of conflicts generated.

In these situations there is often not a right or wrong reason and the two reviewers or a third reviewer will need to come to a final decision about which reason should be recorded.

08

Inter-rater reliability

Inter-rater reliability (IRR)

Conflicts in title and abstract screening

Reliable, consistent and unbiased selection of studies is important for evaluating the consistency and accuracy of assessment within a systematic or other type of literature review. Inter-rater reliability can be used to ensure that the review processes are objective, and that reproducible decisions are made for inclusion and exclusion of studies based on eligibility criteria.

Inter-rater reliability (IRR) is a measure of the consistency and agreement between two or more reviewers, e.g. in their assessments of studies during screening. It is the degree to which different reviewers produce similar or consistent results when evaluating the same study. The IRR score can help to identify discrepancies in screening, ambiguities that may need addressing through training or explanation, and can support the validity of the review process.

IRR can be reported as the percentage agreement (number of agreement scores/total number of scores). It can be measured using statistical methods such as Cohen's Kappa coefficient, intraclass correlation coefficient (ICC), Gwet's AC1, or Fleiss' kappa; which take into account the number of reviewers, the number of categories or variables being rated, and the level of agreement among the reviewers.

High IRR values indicate that the reviewers are consistent in their judgements, whilst low IRR suggests that they have different interpretations or criteria for evaluating the same study. The Kappa scores range from -1 to 1, where 0 represents agreement by chance and 1 represents 100% agreement between screeners.

IRR score levels

Score	Level of agreement
≤ 0	No agreement
0.01 - 0.20	No to slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 1.00	Perfect agreement

Source: Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–174.

Remember IRR can be used during proportional screening to ensure there is a high level of agreement before moving to single-reviewer screening. Achieving high inter-rater reliability is crucial for ensuring the validity and generalisability of research findings or evaluation results.

A low Kappa score indicates lack of agreement

When the level of agreement between reviewers is lower than chance there are likely to be a higher-than-expected number of conflicts requiring resolution. Kappa scores might be low because of:

- **Lack of clarity or ambiguity in the criteria:** When the eligibility criteria used are unclear or ambiguous, reviewers may have different interpretations and produce inconsistent selections.
- **Differences in judgement, perception or preferences:** Reviewers may have different judgements or perceptions of the criteria being evaluated. Personal biases or preferences can influence their evaluations, leading to inconsistent selections.
- **Inadequate training or lack of experience:** If the reviewers are not adequately trained, or do not have a clear understanding of the protocol or criteria, they may produce inconsistent selections.
- **Complexity of the topic:** If the criteria being evaluated are complex or difficult to evaluate the reviewers may have difficulty producing consistent selections.
- **The 'Kappa paradox':** The prevalence of "yes" or "no" votes can skew the coefficient, e.g. in a systematic review where 90% of studies are excluded, the raw agreement might be high, but Kappa can appear artificially low.

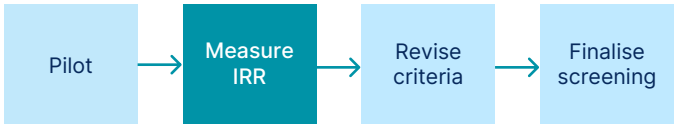
There are several ways to improve inter-rater reliability

By identifying the reasons for low inter-rater reliability, steps can be taken to address them and improve the consistency and accuracy of the study selection.

- **Standardised protocol with clear criteria and definitions:** Ensure that the criteria and definitions used are clear and unambiguous. This can be done through protocols as well as training or discussions. Using instructions, rating scales, and examples of what to look for can be useful.
- **Pilot testing:** Pilot testing can be used to identify any issues with the protocol or criteria before the actual screening process begins. The team may need to discuss disagreements and refine eligibility criteria and possibly recalculate IRR before proceeding with screening of all studies.
- **Training:** Train the review team on how to apply the criteria consistently. This can include piloting exercises for screening, feedback, and discussion.

- **Monitoring:** Monitor the reviewers during the screening process to ensure that they apply the criteria consistently. This can include observing their evaluations, providing feedback, and resolving any disagreements.
- **Blind ratings:** Blinding reviewers can be used to improve inter-rater reliability by preventing individuals from being influenced by the selections made by others.

By using these methods, you can improve the consistency and accuracy of screening, which can lead to more reliable and valid research findings.



09

Automation and machine learning in screening

Automation in screening

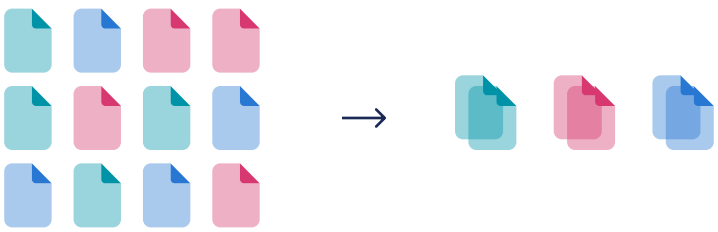
Technology and automation reduces the manual workload and creates efficiencies in the production of systematic reviews. This not only benefits the review community, but also ensures research enters the world faster and provides consumers and end-users with access to the most up-to-date evidence. Often, however, these tools are associated with reduced sensitivity.

The systematic review journey may begin with thousands of potentially relevant studies. Screening them can be time-consuming, laborious and challenging if you are coordinating a large review team.

These challenges have resulted in the development of a number of tools (including Covidence) to support study selection through automation and machine learning. There has been a rapid advancement in automation tools, including text-mining and large language models (LLMs). LLMs are an emerging trend for screening but they require careful validation to avoid missing key studies.

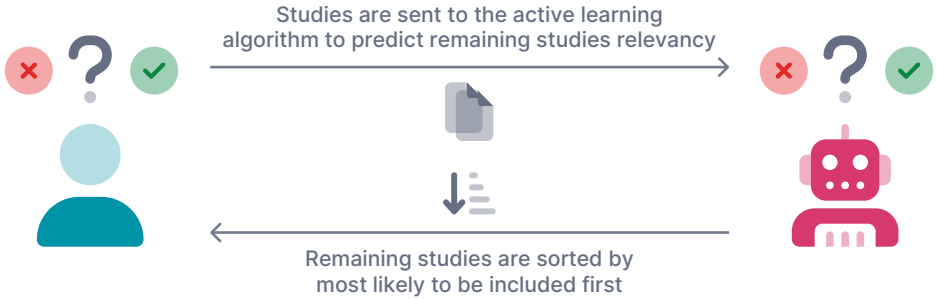
How automation and machine learning can be used in systematic reviews

- **Deduplication** can be done prior to screening using systematic review management and/or reference management software. This reduces the time required for manual identification of duplicates whilst allowing duplication to be verified. Note: automation may not be perfect and it is important to verify suspected duplicates. Some advanced text-mining can flag near-duplicates, for example where titles or authors are slightly different.



- The **Cochrane 'RCT Classifier'** is a machine learning tool that quickly and accurately filters out studies that are not RCTs. The RCT classifier has excellent recall (sensitivity); it can correctly identify a non-RCT with 99% accuracy (Thomas et al. 2021). The tool is available in some systematic review management software tools, such as Covidence, which can also filter out non-RCTs prior to screening.

- **Relevancy** Covidence uses an active learning machine learning model to identify trends in the team's past screening behaviour on the review to determine and display the studies that are most likely to be included first. Human decisions feed back into the model and the more studies you screen, the stronger the system's prediction will be (Miwa, 2014). Some research teams will use relevancy screening to decide when to stop screening, other teams will screen all studies to optimise accuracy and reduce the risk of missing relevant studies.



Covidence sorts studies by most relevant

- **Text mining** and text classification are computational techniques used to manage large volumes of literature efficiently. Text mining is used to automatically extract useful information from a large collection of text documents (e.g manuscripts, abstracts, or reports) and is therefore an effective tool for screening. Text classification is a specific application of text mining where machine learning models or rule-based approaches categorise text into predefined groups (e.g. inclusion/ exclusion criteria, study design, risk of bias).

Reporting the use of automation tools in your review

PRISMA provides recommendations on the reporting of automation tools used in systematic reviews.

Section and Topic	Item #	Checklist item
Selection process	8	Recommendations for reporting in systematic reviews using automation tools in the selection process: • Report how automation tools were integrated within the overall study selection process. • If an externally derived machine learning classifier was applied (e.g. Cochrane RCT Classifier), either to eliminate records or to replace a single screener, include a reference or URL to the version used. If the classifier was used to eliminate records before screening, report the number eliminated in the PRISMA flow diagram as 'Records marked as ineligible by automation tools'. • If an internally derived machine learning classifier was used to assist with the screening process, identify the software/classifier and version, describe how it was used (e.g. to remove records or replace a single screener) and trained (if relevant), and what internal or external validation was done to understand the risk of missed studies or incorrect classifications. • If machine learning algorithms were used to prioritise screening (whereby unscreened records are continually re-ordered based on screening decisions), state the software used and provide details of any screening rules applied.

Tips

- It may be useful to run a pilot on a small subset of studies to evaluate how well the classifier works for that specific topic.
- Consider the extent to which the data used to train a tool are representative of the studies you will screen, and how this might affect the tool's performance. For example, for tools trained on English language text, care is needed for non-English language papers. For tools trained on abstracts, care is needed for conference proceedings.
- If your confidence in the automation tool is low then revert to manual double-screening.

10

PRISMA

PRISMA reporting for screening

Reporting guidelines for screening using PRISMA

You will need to provide a clear summary of the screening process (title and abstract and full text screening) when you have completed the systematic review, and are preparing your report or publication. There is an increased usage of mixed approaches to screening including manual and automated (eliminating references prior to screening, or prioritising references during screening) methods and these will need to be clearly identified and documented.

PRISMA 2020 Statement

Item 8. SELECTION PROCESS: Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.

Your report/publication should provide details on the decisions used to identify if studies met the review inclusion criteria including:

- The number of reviewers who screened each reference (title and abstract) and full text article.
- Whether or not the reviewers screened independently (blinded to other reviewers' decisions) at title and abstract and full text review stages.
- The process used to resolve conflicts between reviewers (consensus or third party).
- Any process that might have been used to contact study investigators for further information.
- Any processes used to translate articles into another language.
- Usage of automation tools during the study selection process (report if decisions were based solely on machine assessments, or verified human decisions). Detail any classifiers (e.g Cochrane RCT classifier) that were used and when they were used in the review process. Use the PRISMA flow diagram to identify 'Records marked as ineligible by automation tools' if classifiers have been used prior to the start of screening.
- Details of any internally derived machine learning classifier, including the version and how it was used and trained (if relevant).
- The software used to prioritise screening using machine learning algorithms and the details of any stopping or screening rules that have been applied.
- Details of crowdsourcing platform/s (if applicable) and how it/they were used to screen records.

Example

After deduplication, we applied Cochrane's randomised controlled trial (RCT) machine learning classifier (Thomas 2021). References identified as being unlikely to be RCTs were removed from further consideration prior to the start of screening. The references were verified by a single reviewer (HH). If an error in selection was identified, the reference was returned to screening.

The remaining references were then screened by two independent reviewers (MM, GG) using Covidence systematic review management software. Conflicts were resolved by consensus and a third reviewer (HH) was used as an adjudicator when conflict resolution could not be reached. Google Translate was used when we found non-English articles to determine eligibility. Citations that did not meet the inclusion criteria were excluded and the reason for exclusion was recorded at the full-text screening

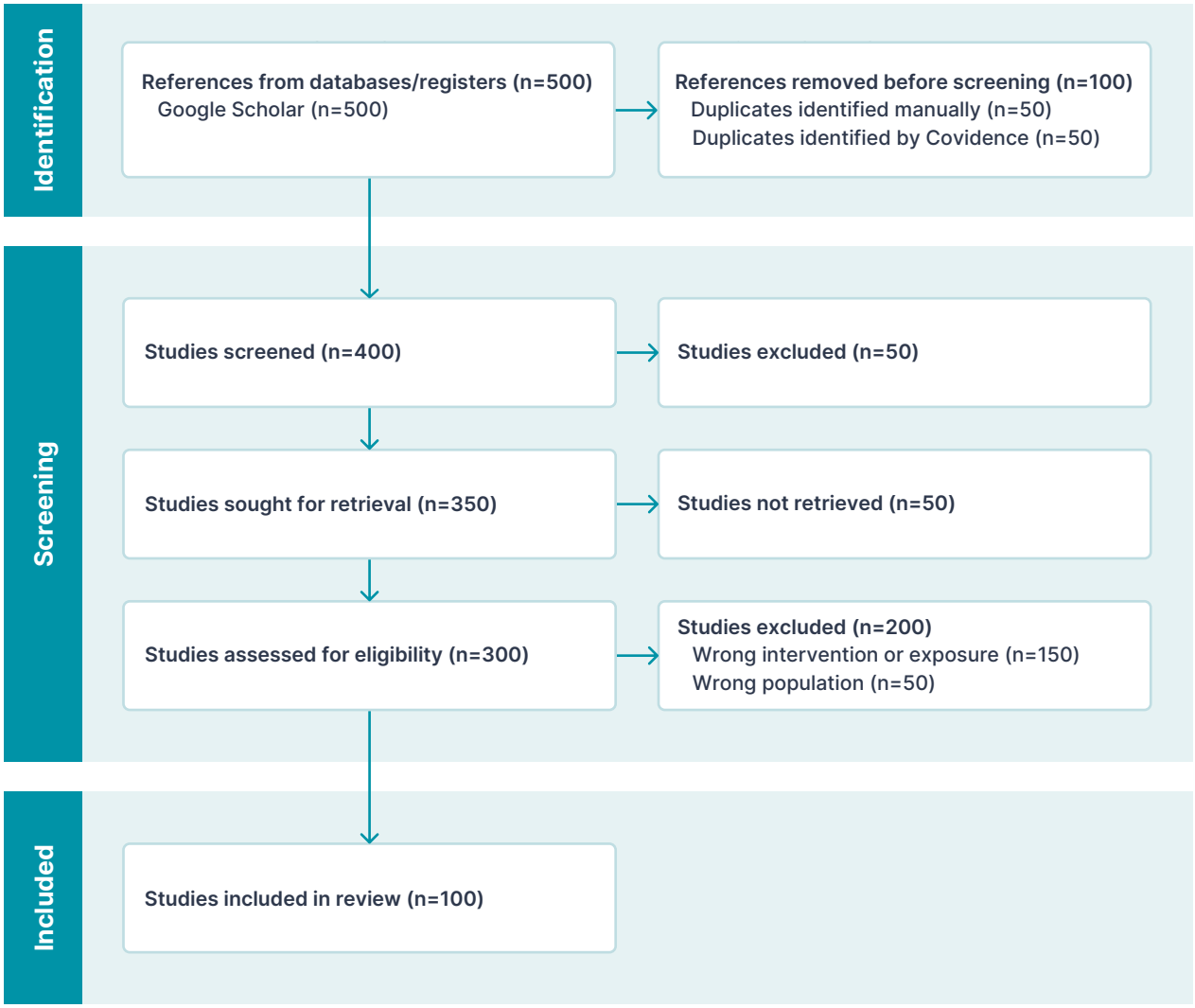
Reporting the flow of studies using PRISMA

During the selection process it is important to track the number of references and subsequently the number of studies so that a flow diagram can be constructed. This provides transparency and enables reproducibility of your methods. This diagram is often referred to as the PRISMA flow diagram.

The PRISMA flow diagram should identify the total number of references/studies:

- retrieved via searching
- identified as duplicates (manual/automatically identified) and other references removed before screening as deemed ineligible
- screened at title and abstract stage
- classified as irrelevant
- screened at full text stage
- excluded and reason for exclusion
- identified for inclusion.

The decisions and reasons for exclusion can be tracked using specialist systematic review software (e.g Covidence), reference management software, a simple document or a spreadsheet. Some software such as Covidence will also track the sources (databases) in PRISMA, when the information is provided.


Identification
References from databases/registers (n=500)
Google Scholar (n=500)
References removed before screening (n=100)
Duplicates identified manually (n=50)
Duplicates identified by Covidence (n=50)
Screening
Studies screened (n=400)
Studies excluded (n=50)
Studies sought for retrieval (n=350)
Studies not retrieved (n=50)
Studies assessed for eligibility (n=300)
Studies excluded (n=200)
Wrong intervention or exposure (n=150)
Wrong population (n=50)
Included
Studies included in review (n=100)

11

Team management

Meeting with the team

The review team should meet throughout the systematic review to ensure the screening process is rigorous, transparent, and consistent. Meetings provide opportunities for collaboration, troubleshooting, and decision-making at critical stages of the review.

- **Protocol development and revision meetings** - These meetings allow the team to agree on eligibility criteria and to allocate roles for screening and conflict resolution at title and abstract and full text stages. If any protocol revisions need to be made that relate to screening the review team should be in agreement and all changes should be documented.
- **Training sessions** - It is important that all team members involved in the screening process understand the process and the eligibility criteria. Training is, therefore, a key step to ensure that the review runs smoothly. It's worth taking the time to train the screeners by piloting some of the studies. It will certainly save time overall and will reduce the number of screening conflicts.
- **Conflict resolution meetings** - The screeners may need to meet to discuss conflicts that can be resolved by consensus. These meetings can be online or face-to-face. They ensure that there is a justified and agreed rationale for including or excluding studies from the review.
- **Progress tracking** - It's worth having some dedicated brief check-in meetings to ensure that everything is on track with screening. These can be online and it is a time when the team can check if they are on target. Early identification of issues and addressing challenges will speed up the review process and keep the review on schedule. These might be opportune times to check inter-rater reliability.

The timing between meetings can vary according to which step in the review process the team is at. It is important to ensure a time slot for subsequent meetings and provide procedures for asking and answering questions in the time between the meetings. Identify who on the team is responsible for this task. Problem solving along the way will save time.

Example agenda

Date	
Attendees	
Agenda	1. Progress update. 2. Discuss any issues or questions relating to screening 3. From point 2, is there a need to amend the protocol or screening guidance to the team? 4. If yes to 3, then decide if a. Do you need to log this change for transparency purposes? b. Who will be responsible for making the change? 5. Time and date of next meeting

12

Top 10 tips on screening

Key takeaways

Top 10 tips on screening

1. Establish your research team early in the review process. Assign tasks to reviewers including roles and responsibilities for screening.
2. Document the processes for screening in your review protocol.
3. Make sure that exports from databases are complete or, at a minimum, contain the reference (citation), abstract and Digital Object Identifier (DOI) link.
4. Always check duplicates removed by automation tools for false positives.
5. Have some brief instructions for the reviewers undertaking screening, especially if there are complicated definitions or conditions.
6. Pilot screening to ensure that the team understands the processes and eligibility criteria, and that any classifiers/automation tools work effectively.
7. Have the eligibility criteria easily accessible to the reviewers who are screening.
8. Retrieve the associated full text PDF or document before you start full text screening of a study.
9. Set up a hierarchy for the reasons for exclusion to minimise conflicts.
10. If your confidence in the automation tool is low then revert to manual double-screening.

Did you know that we have other resources available on our website?

Visit our blog page for insights, announcements, and product updates: www.covidence.com/blog

Already working on a review using Covidence?

Visit our Knowledge Base for a step by step guide on screening within the Covidence platform: https://support.covidence.org/help/screening

Connect with us

Find out more about how we are helping institutions worldwide empower their researchers. Visit www.covidence.org and join our growing social media community.

 LinkedIn

 X (formerly Twitter)

 YouTube

covidence