Chapter 12 **Systematic Reviews**

Systematic review is a method for summarizing literature by collecting and criticizing multiple research studies. A systematic review is a method for qualitative evidence synthesis and is regarded as a source for evidence-based practice.

Traditional literature reviews

Literature reviews do not follow a systematic approach in the selection, appraisal, presentation of clinical studies. Therefore, bias is common when selecting and presenting information in traditional reviews. Additionally, literature reviews do not follow a prespecified protocol making them non-reproducible and merely reflecting the opinion of the review authors.

Systematic reviews

Unlike traditional reviews, systematic reviews must follow a prespecified methodology protocol including details of the literature search and criteria of the information source, criteria for study selection, and appraisal, and the methods of handling and presentation of data. Therefore, systematic reviews involve less bias than traditional reviews, making them, when well-conducted, a valuable source for evidence-based practice. Systematic reviews are reproducible, making the procedure of evidence synthesis more transparent and allow updating evidence when new information becomes available.

Reviews of Cochrane Collaboration

Cochrane collaboration[4] is a network of healthcare professionals, researchers, and statisticians who perform high-quality, state-of-the-art systematic reviews for evidence-based practice. Systematic reviews performed by Cochrane collaboration are usually away from financial interests with industry. Cross-sectional evaluations showed that non-Cochrane review sometimes overestimates treatment effect when compared with similar reviews of the Cochrane collaboration.

Registration of Systematic Reviews

The PRISMA statement[5] recommended that prospective registration of systematic reviews should be a requirement for their publication in medical journals.

---

[4] https://www.cochrane.org/
[5] http://www.prisma-statement.org/

Subsequently, the center for reviews and dissemination has built the PROSPERO database[6] for prospective international register of systematic reviews.

Importance of Systematic Reviews

Clinicians and researchers should not rely on individual studies. Even well-conducted studies can give incorrect estimates. In 2005, Ioannidis et al. found that 32% of large studies (cited more than 1000 times) presented contradictory results that were proved later to be incorrect.

Hundreds of research papers are published every day in worldwide medical journals. The magnitude of published literature in each field is rapidly increasing beyond the capacity of individual researchers and physicians highlighting the need for a well-structured summary of the literature collecting, criticizing, and presenting data from multiple studies.

Additionally, in a clinical setting, the limited time of physicians warrants a trusted summary of the recent clinical evidence. Because they act as a less biased source for evidence-based practice, systematic reviews are rapidly replacing traditional reviews.

Systematic reviews and meta-analysis might have a lifesaving protentional, especially when a small but clinically significant effect estimate is not clear in individual studies — for example, (1) Sudden Infant Death Syndrome and (2) Streptokinase for myocardial infarction.

Types of systematic reviews

Systematic reviews can be classified according to the design of included studies into three categories:

(1) Systematic reviews of observational studies

This type of reviews includes observational studies which can be either (i) cross-sectional studies (to estimate the prevalence of a condition) or (ii) case-control and cohort studies to estimate the effect of exposure to a certain risk factor on the clinical outcome.

(2) Systematic reviews of diagnostic test accuracy studies (DTA reviews)

This type of reviews aims at the evaluation of the accuracy of diagnostic tests. Such reviews are based on data from diagnostic test accuracy studies (DTA studies) where a group of the population is screened by the new test, and the ability of this test to detect positive and negative cases is presented in comparison to the standard

[6] https://www.crd.york.ac.uk/prospero/

diagnosis (index test). As we explained earlier (Chapter 12) that multiple parameters are calculated to represent the accuracy of the test. Some of these parameters are empirically based on the diagnostic accuracy of the test only while other parameters might be influenced by the magnitude of the disease in the underlying population (i.e., predictive values). Given that DTA studies are likely to be heterogenous in terms of the reference test and the source of the study population, DTA reviews should be approached cautiously. Guidelines for performing a DTA review is not explained in this book. However, the Cochrane Handbook of DTA review is a recommended source for further information.

(3) Systematic reviews of interventional studies (clinical trials)

The most important type of systematic reviews is concerned with estimating the safety, efficacy, and cost-effectiveness of treatments. Such reviews are usually based on data from clinical trials where participants are allocated to receive either the experimental or control treatment.

Steps of the systematic review

1- Defining the review question

2- Defining the eligibility criteria

3- Searching the literature

4- Screening of relevant records

5- Quality assessment of included studies

6- Data extraction from included studies

7- Evidence synthesis and data presentation

Defining the review question

The review question of systematic reviews must be specific and supported by clear eligibility criteria. Unlike traditional reviews, systematic reviews usually cover a narrow scope. A systematic review of intervention should satisfy the PICO domains (population, intervention, comparator, and outcome).

For example, Safety and Efficacy of titanium elastic nails compared to spic cast for children with femoral fractures

Defining the eligibility criteria

Review authors should specify clear eligibility criteria based on which relevant studies are included in the evidence synthesis process.

Searching the literature

The term "medical literature" includes medical books, electronic databases, medical journals, and conference proceedings. To run a systematic review, review authors should run a comprehensive literature search to identify all relevant studies. Commonly used databases are PubMed, Scopus, ISI, Embase, EBSCO, Google Scholar, and Embase. The search strategy of systematic reviews should by sensitive rather than specific to ensure including most of the relevant reports and avoid missing potential data. Additionally, searching clinical trial registries (e.g., www.clinicaltrials.gov) is important to reduce publication bias by identifying unpublished studies.

Publication bias

The condition that positive results are more likely to be published than negative results in publication bias. This differential selection of work to be published will introduce bias to medical literature. Common reasons for publication bias are: (1) authors are not more motivated to report and publish positive results, (2) journal editors are more likely to accept papers reporting positive results, and (3) funding agencies are more likely to sponsor the publication of positive results. Additionally, some funding companies might withhold the publication of negative results.
As an effort to reduce publication bias in systematic reviews, the review authors should search several indexing databases and try to identify unpublished studies either by asking experts or searching clinical trial registries.

Screening of relevant records

Following a literature search, review authors will apply the eligibility criteria on the retrieved citations. Because the literature search might yield thousands of citations, it will not be practical to screen the full-text articles of all these citations. Therefore, we recommend that screening is performed at least on two levels. The first step is to screen titles alone or titles and abstracts. Although the title and abstract alone cannot confirm the inclusion of an article in the review, such initial screening is mainly important in excluding irrelevant articles (Rule OUT > Rule IN).

Remove duplicates before the screening

In case that the review authors retrieved citations from multiple sources, it is important that they remove duplicate citations before starting the screening process to save time by avoiding screening the same citations again. The reason is that there is considerable overlap between indexing databases, especially when using the same keywords. In a survey of systematic review authors, Endnote was the most commonly-used software used for removing duplicates among retrieved citations. However, the process of removing duplicates is not 100% accurate since multiple indexing database might vary in the meta-data of the same article. Also,

the indexing meta-data are not deposited on all indexing databases simultaneously. As a result, the issue number, volume, and year might appear on a database and not on the other. Also, the meta-data of one database might include first and last name of the author while another database reports the initials of middle names. However, such situations are very rare, and most of the duplicates are successfully omitted by bibliographic management software as Endnote.

The first level of screening: Title/Abstract Screening

Hereby, we mention four different techniques for title/abstract screening, which is the first level of records screening.

Screening method (1) Hand screening

Hand screening is a traditional method of records screening. In this technique, review authors print the retrieved citations align with the exclusion criteria. Each citation is then marked as (in) or (out) with a specific code referring to the reason for exclusion.

Screening method (2) rating citations on the software

Another method for screening is to use the rating option in the bibliographic management software. In this technique, all citations are imported together to the bibliographic management software, and each citation is rated according to its eligibility for inclusion in the review (for example 5 starts =included, 1 star = excluded, 3 stars = maybe).

Screening method (3) Excel/Google Sheets

Many review authors are exporting citations from bibliographic management software in the form of spreadsheets on Microsoft Excel or Google Drive, which allows for sharing the sheet with multiple reviews. Each reviewer can read the title or title/abstract then rate each record as (in, out, or maybe) in the corresponding row. This method allows multiple reviewers to screen the same citations on the same file. However, this is not an independent multiple reviewers' screening.

Screening method (4) Semi-automated tools

Recently, semi-automated abstract screening tools have revolutionized the screening process of systematic review work, allowing for multiple review authors to screen records independently from each other. Common examples for these tools are Abstrakr, Covidence, and Rayyan. In these tools, the authors export the citations file from the bibliographic management software (i.e., Endnote) then upload the citation file to the semi-automated screening website. Then, the reviewer can invite other reviewers in his team via emails. Other reviewers can join and participate in the

screening process. The user interface of these website will show the title/abstract and three decisions beside each abstract (Yes, No, May be), once the author click on the decision, the abstract disappears, and another one is shown and so on. This process saves time during the screening process. And finally, the review authors will be able to see the citations on which they disagreed, the citations that they are unsure about (maybe), and the citations that they agreed on their inclusion.

The second level of screening: Full-text Screening

Following title/abstract screening, you will have a considerable number of citations that deemed relevant to your review. The next level of screening is full-text screening. In this step, you download the full-text articles of the relevant records (that were included in the title/abstract screening). Then, review authors read the articles carefully to make sure that these articles fully satisfy the inclusion criteria.

Whatever the method you will use for your review work, it is important that each step in the systematic review be performed at least twice by two independent persons starting from the screening phase. This is important to minimize the possibility of errors. Multiple authors might have different views regarding the eligibility of an article for the review. In such a case, the review authors should resolve the disagreement by discussion.

PRISMA flow diagram

The study selection process is usually represented as the PRISMA flow diagram showing the number of citations retrieved from various sources and the flow of selection through duplicates removal, title/abstract screening, then full-text screening.

Reporting reasons for exclusion

According to recommendations of the PRISMA statement, review authors are encouraged to report the reasons for the exclusion of articles that were omitted during the full-text screening. This can be reported in the PRISMA flow chart, as well.

Automatic Screening; is it possible?

Recently, many researchers are trying to develop software that performs automatic screening. However, till the moment, the accuracy of these algorithms has not exceeded 95% and therefore, are not used now. These techniques rely on bioinformatic methods as neural networks and machine learning to teach the software to select relevant citations to the specified eligibility criteria. We believe that within a few years, future development into this area will yield more accurate software and allow for faster and accurate, fully-automated, screening.

Assessing the quality of included studies

Given that systematic review is the process of gathering and collecting results of multiple studies together, the methodological quality of the included studies is an important concern. The quality of the synthesized evidence from the review is mainly dependent on three factors: (1) type of included studies, (2) number of included studies, and (3) quality of included studies (For further information about the quality of evidence, you can read about the GRADE approach).

It is important to assess that quality of all studies included in the systematic review irrespective of their design, sample size, place, and reported outcomes.

Study Design	Quality assessment tool
Randomized Controlled Trials	Cochrane ROB tool
Non-randomized studies	ACROBAT NRSI & ROBINS-II
Observational studies (Cohort, Case-Control, and Cross-sectional studies)	Newcastle Ottawa Scale (NOS scale)
Diagnostic test accuracy studies	QUADAS-II

Cochrane Risk of Bias assessment tool

In this section, we are focusing on explaining the quality assessment of randomized controlled trials using the Cochrane Risk of Bias assessment tool. In this checklist, the review authors examine the ROB in five domains: (1) selection bias, (2) performance bias, (3) detection bias, (4) attrition bias, (5) reporting bias.

1. *Selection bias*

Selection bias is caused by differences in the baseline characteristics of the study group. This difference might be attributed to weak methods of sequence generation or bias introduced during the allocation to the treatment group. The random sequence generation and patient allocation processes are the two steps where selection bias might be introduced. Therefore, to examine the risk of selection bias, we revise the methods used to generate the random sequence and the methods of allocating patients to the treatment groups "allocation concealment."

2. *Performance bias*

Performance bias is caused by the difference between the two groups in the level of care, physician treatment, or exposure to risk factors. Blinding of patients and study personnel is important to reduce the risk of performance bias in randomized controlled trials.

3. *Detection bias*

Detection bias is caused by the difference between the two groups in the outcome measurement. Outcome assessors might unintentionally overestimate or underestimate patient outcomes during the assessment in the experimental or control groups. Blinding of outcome assessors is important to reduce the risk of detection bias.

4. *Attrition bias*

Attrition bias refers to the difference between the two groups in withdrawals. Since patients were randomly allocated to the treatment groups, these groups are balanced for many known and unknown confounders. The presence of patients in their groups maintains this balance in confounders. Therefore, a substantial loss of patients from the study or a difference between the two groups in withdrawals will introduce bias to the data and allowing for confounding variables to influence the study outcomes.

5. *Reporting bias*

Reporting bias refers to a condition where the authors tend to report some outcomes and hide other outcomes. This might be in the form of (1) reporting statistically significant results and not reporting non-significant outcomes, (2) measuring several outcomes and reporting favorable outcomes only, (3) reporting the planned outcomes using an assessment tools/scores other than the planned per protocol.

6. *Other biases*

The review authors might judge that a study suffers from another source of bias (e.g., the study stopped early due to a data-dependent problem or baseline imbalance between the two groups).

For further information and examples for situations where these biases are high, low, or unclear, you can read on Chapter 8 – Part 2 of the Cochrane Handbook of Systematic Review and Meta-analysis of interventional studies (Edition 2011 accessed at http://handbook-5-1.cochrane.org/).

Data extraction

In this step, the review authors extract all important data from the included studies. Usually extracted data include: (1) study design characteristics: date, setting, country, study design, sample size, groups … etc.; (2) baseline characteristics/demographics of the studied population; (3) quality assessment domains: method of sequence generation, allocation method, blinding, data completion, outcome reporting, … etc.; and (4) the study outcomes. Many errors might occur during the data extraction

process. Therefore, multiple reviewers extracting data of the same studies are recommended. Data can be extracted to an online form (built on a survey builder website) or an Excel sheet.

For continuous outcomes††

	Group (1)			Group (2)		
Study ID†	Mean*	SD**	N***	Mean	SD	N
Study 1						
Study 2						

† Study ID = the Last name of the first author + Year of Publication
* Mean or Mean Change in case of pre/post assessment
** SD: Standard deviation of the mean change
*** N: The sample size of the group

For dichotomous outcomes††

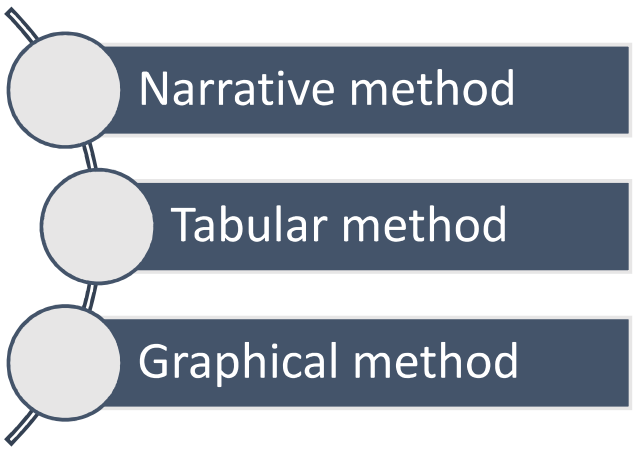
	Group (1)		Group (2)	
Study ID	*No of events*	*N*	*No of events*	*N*
Study 1				
Study 2				

For generic effect size††

Study ID	*ES*	*SE*
Study 1		
Study 2		
Study 3		

†† These tables are for summarization purpose only. The tables indicate the data items required for Review Manager software in order to conduct a meta-analysis. Several other statistical software are also possible with different input forms. For simplification, we present the tables in the format of RevMan.

Presenting the findings



There are many methods to present the results of the systematic review. The narrative method might be used to discuss the evidence and highlight the potential biases in each study. Tabular methods is a better presentation that helps the authors organize the studies according to their design, main findings, and their potential biases. The tabular method is more helpful in the case of evidence derived from several studies. Another method of displaying the systematic review findings is the

graphical method representing the effect estimates reported by the included studies. This graphical presentation is common in Meta-analysis, where the authors employ quantitative analysis methods to pool the effect estimates of multiple studies together. Further information about the methodology of Meta-analysis and their graphical presentation will be discussed in Part II of this handbook.