Meta-Analysis

9

Marco Ceresoli, Fikri M. Abu-Zidan, and Federico Coccolini

9.1 Introduction

Meta-analysis is one of the cornerstones of evidence-based medicine. A meta-analysis is a statistical method allowing to combine the results of two or more studies, giving a pooled estimate result as much closer as possible to the truth, trying to minimize errors. Moreover, the meta-analysis allows to identify differences among the results of the included studies [1].

The rationale to perform meta-analysis is the possibility to collect the results of all the existing studies on a topic and to combine them in a more precise and powerful statistical analysis, based on a higher sample size.

Several types of research data can be analyzed using meta-analysis like comparing an intervention versus another intervention or multiple interventions (randomized controlled studies or case–control studies), results of diagnostic studies, and prognostic data.

Generally meta-analysis is used to combine results of randomized controlled trials, giving the highest level of evidence available, according to the principles of the evidence-based medicine; however, since meta-analysis is only a statistical method,

M. Ceresoli (✉)
General and Emergency Surgery Department, School of Medicine and Surgery,
Milano-Bicocca University, Monza, Italy
e-mail: marco.ceresoli@unimib.it

F. M. Abu-Zidan
The Research Office, College of Medicine and Health Science,
United Arab Emirates University, Al Ain, United Arab Emirates

F. Coccolini
Department of General, Emergency and Trauma Surgery, Pisa University Hospital, Pisa, Italy

it can also be used to combine results of non-randomized studies: in that case the level of evidence of the obtained results is lower. The present chapter describes the fundamental steps needed to (1) perform a meta-analysis and to (2) critically appraise a meta-analysis study.

9.2   The Question

The first step is the definition of the question: This is the fundamental node. The question should follow the PICO model, according to the principles of evidence-based medicine [2]: This stands for (1) P: patients/population, who are the patients or population that you will study? (2) I: intervention: what is the intervention that you are studying? (3) C: control: what is your control? (4) O: outcomes, what are your outcome variables?. This acronym reassumes the fundamental characteristics that a good question should have: a clear definition of the *patients/population* (the disease, for example) in which the investigated *intervention* is compared with a defined *control* for a specific *outcome.*

Once the question is well defined, then the further steps will be a systematic search of the literature with retrieval of all eligible studies, the evaluation of the quality of the studies, and data extraction from each included study. The results of the studies can be pooled and the result of the meta-analysis can be demonstrated by a forest plot.

In this chapter we will use a hypothetical meta-analysis and we will follow it through all the steps. Data are completely invented. Our question is the comparison of laparoscopic appendectomy (intervention) versus open appendectomy (control) in adult patients with acute appendicitis (patients) in postoperative complications and operative time (outcomes). The first step will be the systematic review of the literature.

9.3   Systematic Review of the Literature

A systematic review of the literature is an essential prerequisite before performing the meta-analysis. Once the PICO question is clear, a fundamental step is to define which databases and resources to systematically search and the inclusion and exclusion criteria of these studies. Defining the search protocol with the help of a search methodologist (expert librarian) before starting the systematic review helps to be able to reproduce the results. An inaccurate literature review, without a clear protocol that finds all available relevant data, will lead to biased results. This could happen as a result of an inaccurate review or for the inclusion of "cherry-picked" studies to support a personal viewpoint. For example, if we exclude (accidentally or deliberately) some large sample studies with negative results, we may have a pooled estimate effect influenced by our selection bias. The search, exclusion, and selection process should be described in detail and shown in a flow chart diagram, as recommended by the PRISMA guidelines [3] (Fig. 9.1).
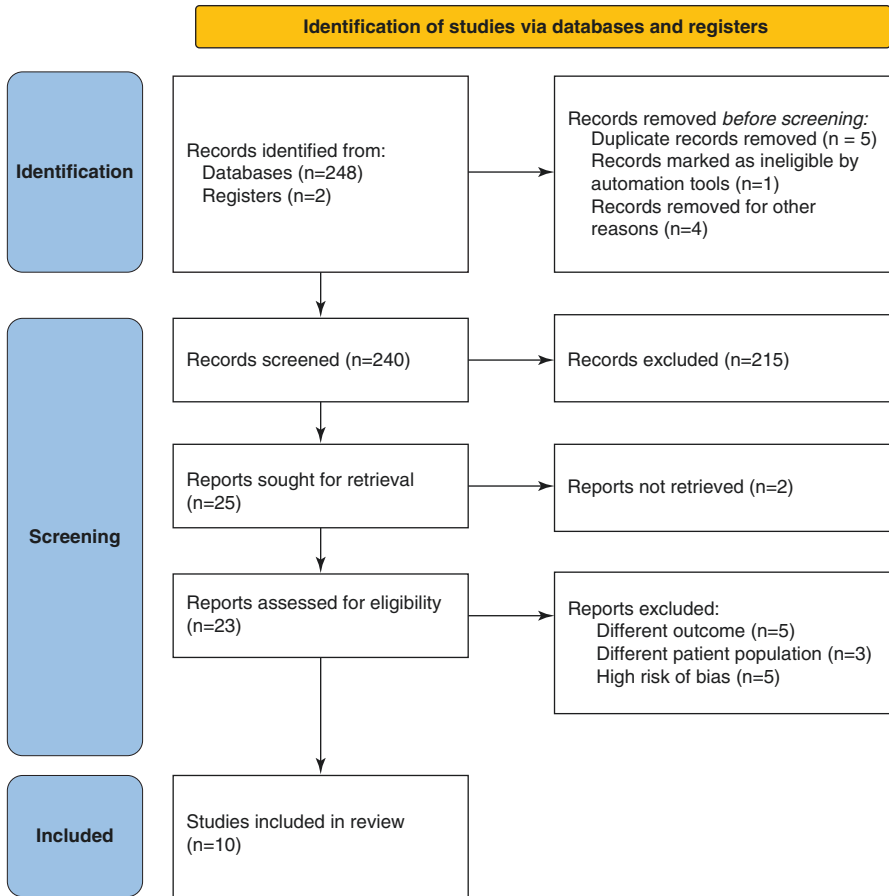
Fig. 9.1 An example of the PRISMA flow diagram

Let us look at our hypothetical example. The flow diagram describes our reviewing process.

The first level contains information about the identifications of studies addressing our topic, based on the criteria adopted and described in the methods section.

The first box describes the number of retrieved records, and the right lateral box contains the number of excluded articles before the screening: we retrieved a total of 250 studies and 10 were initially excluded because duplicate records or other reasons (for example, a study written not in English).

The second level contains information about the screening process and its steps.

The first box describes the first screening (generally made with title and abstract analysis: **Titles are first screened and then those of interest have their abstracts screened**) with the indication of the number of excluded studies and reports indicated in the lateral box: among the remaining 240 studies after a screening of title and abstract 215 were excluded. Two more records were not available giving a total

of 23 studies. The final step of the screening process consists of the assessment for eligibility of the retrieved studies: this process needs an accurate evaluation of the full text of each study; if a study is excluded, we must indicate the reasons for the exclusion. Among the remaining 23 studies we excluded another 13 studies according to the chosen criteria and we indicated the reasons for the exclusion in the lateral box (and in the results section of the meta-analysis). Finally, we have the remaining 10 studies that will be included in the analysis.

9.4   Meta-analysis Appropriateness: Study Inclusion

Another important requirement for a meta-analysis is the absence of considerable clinical or methodological heterogeneity among the selected studies, i.e. the similarity of study design, treatments, and outcomes. Ideally all included studies must have the same design, the same treatment investigated in the same patient's population, and the same endpoint.

There are no statistical tests that could assess and measure clinical heterogeneity and great attention should be given to its description: too precise and narrow inclusion criteria will reduce heterogeneity to the minimum but at the same time they may lead to exclusion of some important studies; conversely, too permissive inclusion criteria will lead to a greater number of included studies but also to a higher clinical or methodological heterogeneity with possible biased results. In case of great clinical heterogeneity, a meta-analysis will not be appropriate. Inclusion and exclusion criteria (on which heterogeneity depends) should be accurately described; You should give great attention to this section when reading a meta-analysis!

Here are some examples of clinical heterogeneity not appropriate for study inclusion:

- the inclusion of a study comparing laparoscopic appendectomy (intervention) versus robotic appendectomy in patients with acute appendicitis when other studies have open appendectomy as the control group.
- the inclusion of a study comparing laparoscopic appendectomy (intervention) versus open appendectomy (control) in only pediatric patients with acute appendicitis (different population) when the other studies evaluate adults.
- the inclusion of a retrospective study comparing laparoscopic appendectomy (intervention) versus open appendectomy (control) in patients with acute appendicitis (population) when the other studies are randomized trials.

9.5   Study Quality Assessment and the Risk of Bias

During the process of studies' evaluation and inclusion in the meta-analysis, it is very important to assess the study quality and the possible risk of bias (see Chap. 4). The presence of bias may under- or overestimate the value of the outcome. Since

the conclusions and the interpretation of the results of a meta-analysis depend on the results of the included studies, the presence of biased results of a single included study especially with large sample may lead to misleading conclusions. Therefore, for each included study, the possible presence of biases should be carefully assessed and described. Several tools and scales have been developed for this purpose.

For **randomized trials**, the Cochrane collaboration developed a specific tool for bias risk assessment [1]. This tool evaluates six specific domains containing all possible sources of biases and evaluates the risk of bias in three levels: low risk, some concerns, and high risk.

The six domains are:

- Bias arising from the randomization process: This domain evaluates if the allocation sequence is random and adequately concealed and if there are differences between the characteristics of the randomized groups.
- Bias due to deviations from intended interventions: This domain evaluates if participants are aware of their assigned intervention during the trial and if investigators are aware of participants' assigned intervention (study blinding).
- Bias due to missing outcome data: This domain evaluates if data for this outcome were available for all, or nearly all, participants who were randomized.
- Bias in measurement of the outcome: This domain evaluates the appropriateness of the method of measuring the outcome in the study and between the groups.
- Bias in selection of the reported result: This domain evaluates if the trial was analyzed in accordance with a pre-specified plan and there is no evidence of selection of the results.
- Overall risk of bias: This domain contains a summary of the risk of bias given by the review's authors (at least two different) on the base of the risk assessed in the previous five domains.

The risk of bias should also be graphically depicted with the dedicated Cochrane tool (Fig. 9.2).

For non-randomized studies other qualitative scales have been developed to assess the potential risk of bias. For surgical non-randomized studies one of the proposed scales is the MINORS (Methodological Index for NOn-Randomized Studies) which evaluates 12 items assessing all domains and possible source of biases [4].

Among all the possible biases the publication bias can be graphically depicted and evaluated with a specific graph: the funnel plot. The funnel plot is a scatter plot in which each dot represents a study, and it is allocated in the plot based on the study results (effect size on x axis) and the study precision (the inverse standard error or the number of cases, on y axis). If there is no publication bias the graph will represent an inverse funnel; in case a publication bias is present, the distribution of the dots will be skewed and asymmetric (Fig. 9.3).

Fig. 9.2 Summary of risk of bias. The red dots indicate high risk of bias, green dots indicate low risk of bias, and white spaces indicated uncertain risk of bias
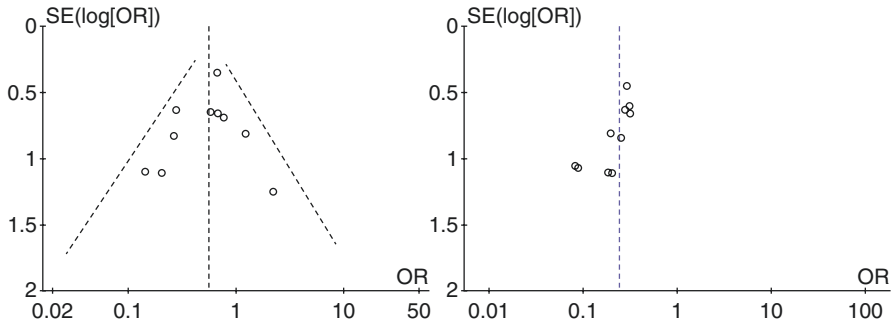
Fig. 9.3 Two example of funnel plots: on the left the distribution of studies is symmetrical (no publication bias); on the right the distribution of the studies is skewed (possible publication bias)

9.6   Results: Effect Measure

The main result of a meta-analysis is expressed with the effect measure. The effect measure is a statistical construct that compares outcome data between two intervention groups (intervention vs. control). The effect measure depends mostly on the type of data analyzed. Two general groups of effect measures exist: the ratio measures (for dichotomous outcomes) and the difference measures (for continuous outcomes).

According to the type of the data these are the most commonly adopted effect measures.

9.6.1   Binary Outcomes/Dichotomous Data

- Risk ratio (RR): It is the ratio between the risk of an event in the two different groups X and Y (see Chap. 8); it can be a number between 0 and infinite where 1 is the no effect value (same risk in the two different groups). When the risk of the event complication is higher in the laparoscopic appendectomy group than open appendectomy group, the RR will have value >1; on the contrary when the risk of the event is higher open appendectomy group than laparoscopic appendectomy group, the OR will have a value between 0 and 0.99. This is the preferred measure for randomized studies' outcomes. A RR = 1.56 should be interpreted as 56% higher risk of complications in laparoscopic appendectomy group compared with open appendectomy; RR = 0.56 should be interpreted as a 44% reduction of complication in laparoscopic appendectomy group.
- Odds ratio (OR): Similarly, to the RR this measure is the ratio between the odds of the event in the two compared groups (see Chap. 8). The measure is a number between 0 and infinity, where the value 1 corresponds to no effect (same odds in the two groups). When the probability of complication is higher in the laparoscopic appendectomy group than open appendectomy group, the

OR will have value >1; on the contrary when the probability of the event is higher in open appendectomy group, the OR will have a value between 0 and 0.99. Odds ratio should be adopted in meta-analysis of case–control studies. Differently from RR an OR = 1.56 does not correspond to a 56% increase in the risk! Its value could approximate the RR only when the frequency of the event is less than 10%.

9.6.2 Continuous Data (Also Scale Data or Counts of Events)

- Mean difference (MD): It measures the absolute difference between the mean values of two compared group, giving a numeric value that represents the pooled difference. The effect size provides information expressed as a clinical unit (for example, the mean difference of operating time, in minutes; Fig. 9.5). It is appropriate when all study results are expressed in the same measurement's unit.
- Standardized mean difference (SMD): When study results are available in different measurement units, continuous results can be meta-analyzed through the standardized mean difference that provides information expressed as statistical units. The standardized mean difference measures the effect on the base of data dispersion and it represents the effect expressed in number of standard deviations (SD) (differently from mean difference that is expressed in clinical unit as minutes, days, or milliliters of blood loss). A SMD of 1.1 represents a variation of 1.1 SD. Generally, the value 0.2 is considered as a small effect, 0.5 as medium, and 0.8 as large effect. This measure is not easy to be interpreted and it is useful in limited cases of surgical studies.

9.7 Results: The Forest Plot

Forest plots are the preferred graphs for reporting the results of meta-analysis. They contain several information about the meta-analysis. In this section we will show the forest plot created by the Cochrane RevMan software, the open-source tool provided by the Cochrane organization for making meta-analysis. Figure 9.4 shows the forest plot containing the results of our hypothetical meta-analysis with the comparison of a dichotomous outcome, morbidity rate, between laparoscopic and open appendectomy.

The forest plot is built as a combination of a table and a graph. On the left side are shown the results of each included study: the first line shows data about "Study A" with the number of events in the experimental and control treatment groups and the respective number of patients in each group. Each study has a "relative" weight in the meta-analysis: this weight is based on the study precision: the narrower is the 95% confidence interval (more precise data, small variance), the higher will be the weight; on the contrary, a study with a wide 95% confidence interval (CI) will have a lower weight. Finally, for each study is represented the effect estimate (the result of the study is represented with the chosen effect measure, along with its 95% CI).
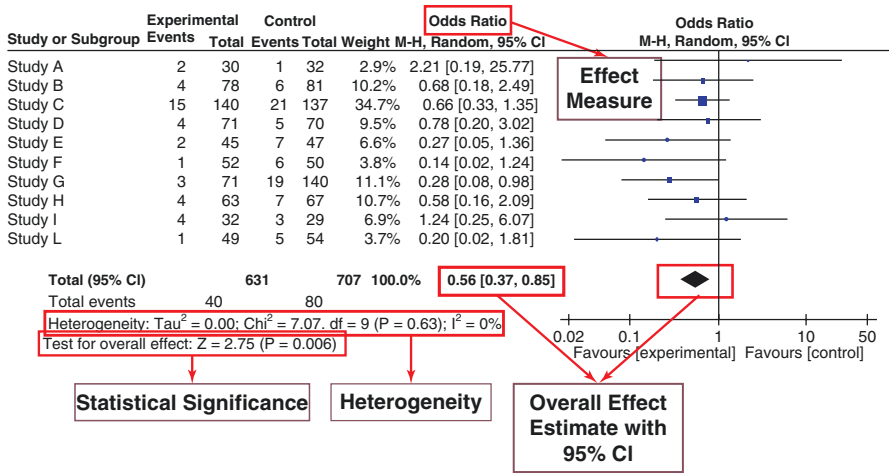
Study or Subgroup	Experimental Events	Total	Control Events	Total	Weight	Odds Ratio M-H, Random, 95% CI
Study A	2	30	1	32	2.9%	2.21 [0.19, 25.77]
Study B	4	78	6	81	10.2%	0.68 [0.18, 2.49]
Study C	15	140	21	137	34.7%	0.66 [0.33, 1.35]
Study D	4	71	5	70	9.5%	0.78 [0.20, 3.02]
Study E	2	45	7	47	6.6%	0.27 [0.05, 1.36]
Study F	1	52	6	50	3.8%	0.14 [0.02, 1.24]
Study G	3	71	19	140	11.1%	0.28 [0.08, 0.98]
Study H	4	63	7	67	10.7%	0.58 [0.16, 2.09]
Study I	4	32	3	29	6.9%	1.24 [0.25, 6.07]
Study L	1	49	5	54	3.7%	0.20 [0.02, 1.81]
Total (95% CI)		631		707	100.0%	0.56 [0.37, 0.85]
Total events	40		80			

Heterogeneity: Tau^2 = 0.00; Chi^2 = 7.07, df = 9 (P = 0.63); I^2 = 0%
Test for overall effect: Z = 2.75 (P = 0.006)

Fig. 9.4 A forest plot showing a comparison of a dichotomous outcome (complications following experimental treatment compared with control)

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Study A	120	18	30	118	35	32	4.7%	2.00 [−11.73, 15.73]
Study B	131	17	78	140	19	81	12.9%	−9.00 [−14.60, −3.40]
Study C	121	27	140	130	28	137	11.5%	−9.00 [−15.48, −2.52]
Study D	121	7	71	131	9	70	17.5%	−10.00 [−12.66, −7.34]
Study E	115	27	45	121	24	47	6.9%	−6.00 [−16.46, 4.46]
Study F	124	37	52	137	28	50	5.3%	−13.00 [−25.70, −0.30]
Study G	119	27	71	120	31	140	9.3%	−1.00 [−9.11, 7.11]
Study H	117	31	63	121	30	67	6.9%	−4.00 [−14.50, 6.50]
Study I	124	9	42	121	14	29	12.6%	3.00 [−2.78, 8.78]
Study L	131	16	49	137	15	54	12.2%	−6.00 [−12.01, 0.01]
Total (95% CI)			641			707	100.0%	−5.69 [−9.12, −2.26]

Heterogeneity: Tau^2 = 15.64; Chi^2 = 22.24, df = 9 (P = 0.008); I^2 = 60%
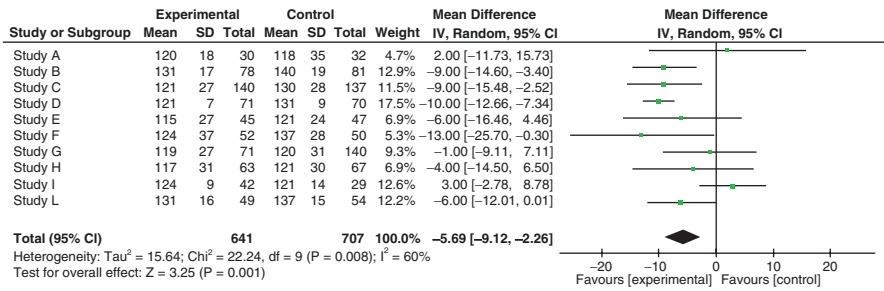Test for overall effect: Z = 3.25 (P = 0.001)

Fig. 9.5 A forest plot showing a comparison of a continuous outcome (operative time in minutes between two groups)

The effect measure is also depicted in the right part of the plot: the effect estimate is shown as a box and its dimension varies according to the study's weight (higher weight has bigger dimensions); the line represents the 95% CI.

The last line of the forest plot shows the results of the meta-analysis: the overall number of events and patients in experimental and control groups and the overall effect estimate. The effect estimate is a pooled estimation of the effect of all included studies, adjusted according to each study's weight. On the right it is shown as a diamond, having a width which represents the 95% CI.

On the right side, where effects are graphically shown, there is a vertical line: this line represents the line of "no effect." This line corresponds to the value "1" when the effect measure is a ratio (odd ratio, risk ratio) and the value "0" when the effect measure is a difference (risk difference, mean difference, standardized mean difference, see Fig. 9.5). The position of the diamond gives a graphical representation of

the meta-analysis results: when the diamond lies entirely to one side of the line, there is a significant difference between the groups (the "no effect" value is not contained in the 95% CI). If the diamond is on the left of the line, the effect measure shows a lower frequency of events in the experimental group (a result favoring the experimental group in case of bad outcome, as complications or deaths, or favoring control group in case of good outcomes as cure, success of the therapy). On the contrary, if the diamond lies in the right of the line the result should be interpreted as favoring control group in case of bad outcomes.

On the bottom line there are information about the statistical heterogeneity (I^2) and the statistical significance of the analysis (test of overall effect Z).

In our hypothetical example, in which we analyzed the effect of laparoscopic appendectomy (experimental) compared with open appendectomy (control) on the complications rate, we included all the ten studies retrieved (from A to L). The overall effect showed a significant reduction of complications with laparoscopic appendectomy with an effect measure expressed as odds ratio of 0.56. This means that laparoscopic appendectomy reduced the complications by approximately 44% compared with open appendectomy. The confidence interval for the point estimates was 0.87–0.85.

Figure 9.5 shows the comparison of a continuous outcome (operative time) between our two chosen surgical interventions. The mean and the standard deviation for experimental and control groups are represented for each study; the weight of each study is calculated based on the data dispersion: higher SD corresponds to a lower weight. The chosen effect measure was the mean difference. The meta-analysis resulted in a significant reduction of operative time of −5.69 min (95% confidence interval −9.12; −2.26). We must notice that, despite statistical significance, the difference between the two treatments is clinically irrelevant (only 5 min difference).

9.8   Results: Heterogeneity

Heterogeneity is a fundamental aspect to be aware of when reading and performing a meta-analysis. It is defined as the presence of differences among studies. There are several kinds of heterogeneity:

- Clinical heterogeneity: A difference in the clinical setting or intervention of the included studies. This should be carefully described. If there was serious heterogeneity, then the meta-analysis may not be appropriate. An example, performing the same interventions (open versus laparoscopic appendectomy) but in different patients' populations (adult patients versus pediatric patients).
- Methodological heterogeneity: A difference in the study design. When present, the meta-analysis could be inappropriate. However, occasionally methodological heterogeneity can be overcome by using the subgroup analysis. An example for that is the presence of randomized and non-randomized studies in the same meta-analysis.

- Statistical heterogeneity: This indicates a difference in the results of the included studies. This may occur because the confidence intervals are not overlapping or because of differences in the direction and the magnitude of the effect of different studies. Statistical heterogeneity of the direction of the effect indicates that the beneficial or harmful effect of the treatment is not similar across the included studies. For example, in Fig. 9.4, in studies A and I the experimental treatment resulted in a harmful effect while in all the other studies had a beneficial effect: this represents a statistical heterogeneity.

Statistical heterogeneity is evaluated using the Chi squared test for heterogeneity with its *p*-value, indicated in the bottom line of the forest plot. A further evaluation is the inconsistency (the measure of incoherence among results), indicated by the I^2. I^2 represents the variation across studies due to heterogeneity. Generally, an I^2 value of less than 40% can be considered as not important, 40–75% as moderate, while more than 75% as substantial.

Heterogeneity conditions the calculation of the meta-analysis results. There are two statistical models for the calculation of the overall estimated effect: the fixed model and the random model. The fixed model is more accurate (narrower CI) but requires an absence of heterogeneity; the random model takes into account statistical heterogeneity and gives more solid results which avoid misinterpretations.

9.9 Interpretation of the Results

A meta-analysis is the result of a very complex and tedious work. The forest plot, that contains all the essential results, should be considered as "the tip of the iceberg" and the interpretation of the results should be a very accurate and cautious. When reading a meta-analysis, we must be familiar with the concept of certainty of the results, defined as the confidence that the true effect is within a particular range or threshold. In other words, certainty is the confidence that the pooled result is true and does not depend on heterogeneity and bias.

The point estimate of the measured effect gives us the direction and the magnitude of the effect. In Fig. 9.4 for example, the experimental treatment leads to a reduction of the outcome (complications) with a measured effect expressed as odd ratio of 0.56. This measure does not alone give us all the information we need to know. One of the most important information is the width of the confidence interval, in which we are 95% confident that the measured effect lies. In our example the confidence interval is between 0.37 and 0.85 giving us a reasonable certainty.

Great attention should be directed towards the difference between clinical and statistical significance: often, a statistically significant result (with a 95% confidence interval that does not contain the "no effect" value or a *p*-value <0.05) is not clinically significant. Figure 9.5 shows that the experimental treatment resulted in a lower operative time with a mean difference of −5.69 min (95% CI −9.12; −2.26). Although statistically significant, 5 min mean difference is clinically not important. Being expert in the studied area is very important to differentiate between clinical

and statistical findings. We should not simply look through the narrow hole of the *p*-value.

The interpretation of the results when there is no significant difference between the two groups raises more difficulties. The absence of significant difference does not allow us to automatically conclude that the two compared treatments are equivalent. In this case, it is very important to differentiate between "true" no effect and uncertainty of the results, based on the evaluation of the width of the CIs.

9.10　Sensitivity Analysis

Since a meta-analysis is mainly a systematic review of the literature, there are several decisions that the researcher must take. Some of these decisions could be arbitrary and not objective. For example, the decision to adopt a numerical value as a cut-off for age, the decision to consider patients who were lost at follow-up as dead, or the decision to include or exclude a study for different reasons. All these elements could influence the results of the meta-analysis. Therefore, they should be analyzed with a sensitivity analysis, to evaluate their role as a possible source of variability. Sensitivity analysis is defined as a repetition of the analysis by changing the included elements or changing the arbitrary or unclear decision criteria. Sensitivity analysis evaluates the robustness of the results of the meta-analysis. The main factors that may implicate a sensitivity analysis are:

- the inclusion and exclusion criteria,
- the clinical or methodological design of studies (source of heterogeneity),
- the model adopted for the analysis,
- the effect measure chosen (for example, fixed effect vs. random effect, odd ratio vs. risk ratio).

Another example of a sensitivity analysis is the repetition of the analysis excluding studies by dimension (generally the exclusion of small studies) or by the presence of heterogeneity.

9.11　Common Mistakes Encountered in Submitted Systematic Review Manuscripts

These are some of the common mistakes we have encountered as reviewers in systematic review articles submitted to surgical journals that may lead to rejection of these papers. Highlighting these errors may help young researchers to avoid them. These errors include:

1. *Mixing between a systematic review, scooping review, and a narrative review*: A narrative review, although searches the literature, has a broad scope and does not follow the strict rules of systematic reviews which have a precise protocol

and search methods. It is subjective, affected by personal opinion and selection bias [5]. A scoping review, similar to systematic review should have a clear methodological protocol to reproduce the results [6]. It differs from a systematic review in two aspects: (1) including a minimum of one search engine, (2) having a broad research question [7], otherwise the methodology is the same.

2. *Unclear or unimportant research question*: It is very important to define an important focused research question. Systematic reviews may take up to 18–24 months of continuous work to be properly performed. Systematic reviews answering the same question will usually give the same answer if they follow the same methodology. Accordingly, it is important to check whether there are similar systematic reviews in the literature that answered the same question so this major effort can be utilized in the proper direction.

3. *Lack of a clear structured protocol*: This protocol should be written to be detailed so as to be followed when performing the study. It should define the search strategy, terms, outcome variables, and methods of statistical analysis.

4. *Lack of search experience*: Systematic reviews depend entirely on the search process. The literature search needs both a subject expert and a search methodologist to be useful. It should have enough technical details that can reproduce the study if done by others. This includes using appropriate truncations like (*) and using synonyms to assure retrieving and covering all core keyword variations and locating all possible evidence. For example, putting words between brackets will only search the exact sequence of the words and spaces and not individual words.

5. *Not properly following the protocol and inclusion exclusion criteria*: Systematic reviews by definition are original articles that have detailed methodology that can be reproduced by any researcher if methods were followed. The subjects of the study are the included articles. The authors should follow exactly the protocol of the study.

6. *Not documenting the search procedure*: This is a common mistake. The authors may really do a systematic review in a specific time using specific search engines and specific terms but do not document them. If not fully documented, the authors will not be able to reproduce the results. It is very important to document each step when doing the search so the PRISMA graph can be accurate and reproducible.

7. *Being too narrow in the search*: Some authors narrow the search without a justification to reduce the effort needed in performing a systematic review. They may narrow the period of the studies, the geographical location, or the search engines. A systematic review needs a minimum of two databases (we recommend at least PUBMED and EMBASE). The more databases are searched, the better the systematic review will be.

8. *Lack of critical appraisal and improper evaluation of the quality of the selected papers:* The authors should evaluate the quality of the studies even if the studies were retrospective. It is advised to have a minimum of two research methodologists who independently critically appraise the selected papers. This is very important to exclude papers being published twice either by increasing sample

size (in which the first should be excluded) or finding dual publication of the same data.

9. *Overusing statistics:* It is very important to know when not to do a meta-analysis. Just to clarify this issue, you cannot mix apples and oranges and count them together. Furthermore, adding combing weak studies or heterogenous studies does not increase the quality of the evidence.

10. *Not acknowledging biases*: The authors should recognize all relevant biases of a study including geographical bias, language bias, search bias, etc. This indicates that the authors were aware of the limitations of their study. It is advised to include this in detail in the limitations section [8].

9.12 Conclusions

Meta-analysis is a statistical technique that allows to combine the results of two or more studies. Meta-analysis cannot exist without a systematic review of the literature. Reading and understanding a meta-analysis is much more complex than looking at the forest plot. A "check-list" for a correct reading and interpretation of these complex studies includes:

- Accurate and precise literature review.
- Precise definition of inclusion and exclusion criteria.
- Description of retrieved studies with reasons for inclusion and/or exclusion.
- Assessment of the study quality and the potential risk of bias.
- Description of heterogeneity (clinical, methodological, and statistical).
- Evaluation of the correct effect measure.
- Assessment of statistical significance vs. clinical significance.

One of the commonest errors for the reader is to concentrate and give attention only to the forest plot drawing conclusions without critically reading the whole study. The robustness of the results should be accurately evaluated (with sensitivity analysis, for example). Even in case of statistically significant results (the diamond in the forest plot does not cross the no effect line), the presence of important heterogeneity could question the certainty of the results, and no definite conclusion can be reached. More specific and detailed description of the meta-analysis methodology can be found in the Cochrane handbook for systematic reviews and meta-analysis [1].

References

1. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions. Oxford: The Cochrane Collboration and Wiley; 2019. https://training.cochrane.org/handbook. Accessed 2 Apr 2022.
2. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations.

BMJ. 2008;336:924–6. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2335261&tool=pmcentrez&rendertype=abstract.

3. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

4. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J Surg. 2003;73:712–6. http://www.ncbi.nlm.nih.gov/pubmed/12956787.

5. Papakostidis C, Giannoudis PV. Systematic reviews and meta-analyses: what are the common pitfalls? Injury. 2022;53:1301–4.

6. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7):467–73.

7. Powell JT, Koelemay MJW. Systematic reviews of the literature are not always either useful or the best way to add to science. EJVES Vasc Forum. 2021;54:2–6.

8. Mohammad A, Branicki F, Abu-Zidan FM. Educational and clinical impact of Advanced Trauma Life Support (ATLS) courses: a systematic review. World J Surg. 2014;38:322–9.