A Practical Guide to Writing (and Understanding) a Scientific Paper: Meta-Analyses

47

Alberto Grassi, Riccardo Compagnoni,
Kristian Samuelsson, Pietro Randelli, Corrado Bait,
and Stefano Zaffagnini

47.1 Introduction

In the era of evidence-based medicine, clinicians are continuously facing the massive production of clinical studies, often with discordant evidence. To facilitate the spread of knowledge, narrative reviews, systematic reviews and meta-analyses represent indispensable tools for summarising the evidence related to a specific topic.

In a *narrative review*, an "expert" summarises the important aspects relating to a topic. It is assumed that this expert will be objective in presenting the pertinent information. Since a specific research question and a systematic search of the evidence are lacking, bias (usually unintentional) is often a problem [19].

Conversely, a *systematic review* represents a higher and unbiased level of evidence, since it assimilates information about a topic or question with more rigour, sophistication and transparency. A systematic review is a formal process for gathering and evaluating literature to answer a specific question, beginning with the posing of the question, the definition of the inclusion and exclusion criteria of trials and extracting the necessary data from each one. Moreover, methodological quality evaluation is usually performed [19].

A *meta-analysis* represents a further step, as it statistically combines the data from clinical

A. Grassi (✉)
Dipartimento Scienze Biomediche e Neuromotorie DIBINEM, Università di Bologna, Bologna, Italy

IIa Clinica Ortopedica e Traumatologica, IRCCS Istituto Ortopedico Rizzoli, Bologna, Italy

SIGASCOT Arthroscopy Committee, Florence, Italy
e-mail: alberto.grassi3@studio.unibo.it

R. Compagnoni
Società Italiana del Ginocchio Artroscopia Sport Cartilagine Tecnologie Ortopediche (SIGASCOT) Arthroscopy Committee, Florence, Italy

1st Department, Azienda Socio Sanitaria Territoriale Centro Specialistico Ortopedico Traumatologico Gaetano Pini-CTO, Milan, Italy

K. Samuelsson
Department of Orthopaedics, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

Department of Orthopaedics, Sahlgrenska University Hospital, Mölndal, Sweden

P. Randelli
1st Department, Azienda Socio Sanitaria Territoriale Centro Specialistico Ortopedico Traumatologico Gaetano Pini-CTO, Milan, Italy

C. Bait
Società Italiana del Ginocchio Artroscopia Sport Cartilagine Tecnologie Ortopediche (SIGASCOT) Arthroscopy Committee, Florence, Italy

Istituto Clinico Villa Aprica, Como, Italy

S. Zaffagnini
IIa Clinica Ortopedica e Traumatologica, IRCCS Istituto Ortopedico Rizzoli, Bologna, Italy

Laboratori di Biomeccanica e Innovazione Tecnologica, Istituto Ortopedico Rizzoli, Bologna, Italy

471

trials, often randomised controlled trials (RCT), obtained from a systematic review of the literature. The reason for combining the data is an attempt to increase the ability to see a difference between two groups, reducing the chance of type-II errors (missing the existence of a true difference), and to increase the precision of the estimated effect by increasing sample size. Meta-analyses are therefore a powerful tool for accumulating and summarising knowledge; however, their use is also controversial, as there are several critical conditions and methodological considerations that could produce misleading conclusions. For this reason, a meta-analysis requires personal judgement and expertise and the implementation of procedures and quality standards to reduce the risk of bias that may influence the results [11, 30] (Fact Box 47.1).

The aim of this review is to provide the reader with the basic knowledge to write and understand a meta-analysis, describing all the methodological steps in its preparation and providing a useful reference for a detailed, in-depth understanding of this complex and controversial field.

This guide has been developed with four different aims:

1. To provide a guide to prepare a meta-analysis for those clinicians approaching the meta-analysis study design
2. To provide a summary and a source of references for those clinicians already familiar with the production of meta-analyses
3. To help the reader understand a meta-analysis and interpret its results
4. To help the journal reviewer identify the critical aspects of meta-analyses

For a more detailed guide, consulting the *Cochrane Handbook for Systematic Reviews of Interventions* [15] and the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)" [25] is warmly recommended.

47.2 How a Meta-Analysis Works: The Concept of Effect Size

To summarise, a classical meta-analysis of RCTs provides a single, overall measurement of the treatment effect, enhancing the clinical interpretation of findings across several studies [17]. This technique can prove especially useful when there are several similar clinical trials with or without consistent outcomes or when there are small- to medium-sized trials with inconclusive results. The statistical calculation behind the results of a meta-analysis is based on the concept of "effect size". Effect size is defined as a quantitative measure of the magnitude of a phenomenon, and in statistics, it is a parameter that reflects the magnitude and direction of the treatment effect for each study. For example, if all the studies in the meta-analysis measure a continuous outcome, such as the Lysholm score after single- or double-bundle ACL reconstruction, the mean difference between the two groups can be used as the effect size and therefore to express the effects of the treatment. The overall effect size derived from the meta-analysis is calculated by combining the effect sizes of the included studies.

The magnitude and direction of the effect size are integrated with the size confidence intervals (CI). CIs are reported as a probability (e.g. 95% confidence interval, 95% CI), providing a range

---

Fact Box 47.1
- **Narrative review**: when an "expert" summarises the important aspects relating to a topic without a systematic presentation of the evidences. It is usually source of unintentional bias.
- **Systematic review**: formal process for gathering and evaluating literature to answer a specific question, beginning with the posing of the question, the definition of the inclusion and exclusion criteria of trials and extracting the necessary data from each one.
- **Meta-analysis**: process to statistically combine the data from clinical trials, often randomised controlled trials (RCT), obtained from a systematic review of the literature.

with an upper and lower boundary that indicates the precision of the treatment estimates of the effect size of the included studies. A wider CI, which may be caused by a small sample size or by imprecision in the measurement (e.g. wide standard deviations in the considered outcome), indicates less precise estimates and could therefore question the application of the results in clinical practice.

47.3   How to Start Your Meta-Analysis Properly

47.3.1 Define a Study Question

The first and most important step in preparing a meta-analysis is to identify an appropriate question to address [26]. According to the *Journal of Bone and Joint Surgery* guidelines (Fact Box 47.2) [36], a meta-analysis should address a question that has not been considered in the previous 5 years, unless the most recent literature has changed dramatically. Furthermore, the question at the centre of a meta-analysis should not have already been answered satisfactorily by the results of multiple well-conducted RCTs. Finally, the question should be addressed to the evaluation of the effects of two different treatments for the same clinical condition, to allow the inclusion of only randomised or quasi-randomised clinical trials. Typical topics of meta-analyses in orthopaedic surgery could be a comparison of clinical results and the failure rate after single- or double-bundle and hamstring or bone-patellar tendon-bone anterior cruciate ligament (ACL) reconstruction, the re-rupture rate and complications after the conservative or surgical treatment of Achilles tendon ruptures, the re-dislocation rate after a brace or repair after primary patellar dislocation, clinical outcomes and alignment after patient-specific instrumentation (PSI) or conventional total knee arthroplasty (TKA), functionality and pain after hyaluronic acid or platelet-rich plasma (PRP) for knee osteoarthritis, to mention just a few.

An unconventional use of a meta-analysis could be employed to pool together the data from single-arm case series studies. Continuous mea-

> **Fact Box 47.2: Journal of Bone and Joint Surgery (JBJS) Requirements to Perform a Meta-analysis** [36]
>
> – JBJS will not accept meta-analyses or systematic reviews on the same topic published within 5 years unless the authors can demonstrate that the literature has dramatically changed.
> – Meta-analyses or systematic reviews will not be accepted if the same (or largely the same) papers are used to arrive at similar conclusions.
> – For meta-analyses in which the authors use statistical methods to combine and summarise results, only summaries of randomised trials will be accepted.
> – Only studies with sufficient homogeneity of inclusion and exclusion criteria will be considered appropriate for meta-analysis.
> – Authors should familiarise themselves with reporting checklists such as the PRISMA to improve the quality of reporting.

surements such as the Lysholm score, or dichotomous variables, such as return to sport [9], failures, or good vs. poor outcomes [10, 28], could also be pooled, to obtain a mean value for a wider population composed of the sum of the populations of single studies. In certain circumstance, with wide and non-heterogeneous populations, statistical calculations could be used to compare the outcomes for different groups [7]; however, the results should be interpreted with extreme caution, since the data derive from single-arm case series and not from RCTs.

47.3.2 Perform an Appropriate Literature Search

The next practical step is to obtain the largest number of studies relevant to the study question [20]. The literature search should usually be performed by two independent investigators using at least

three databases, as there is an incomplete overlap of the results from single databases. The three bibliographic databases generally regarded as being the most important sources to search for the reports of trials are CENTRAL, PubMed and EMBASE (Excerpta Medica Database). The Cochrane Central Register of Controlled Trials (CENTRAL) serves as the most comprehensive source of reports of controlled trials. CENTRAL is published as part of *The Cochrane Library* and its access is free of charge. Medical Literature Analysis and Retrieval System Online (MEDLINE) currently contains over 16 million references to journal articles from the 1950s and onwards. PubMed provides access to a free version of MEDLINE that also includes up-to-date citations not yet indexed for MEDLINE. Lastly, EMBASE currently contains over 12 million records from 1974 and onwards. EMBASE.com is Elsevier's own version of EMBASE which, in addition to the 12 million EMBASE records from 1974 and onwards, also includes over seven million unique records from MEDLINE from 1966 to the present day. Access to EMBASE is only available by subscription.

In addition to the previous three main databases, other sources should be searched, such as national and regional databases, tables of contents of relevant journals (such as the *Journal of Bone and Joint Surgery, Arthroscopy, the American Journal of Sports Medicine, Knee Surgery Sports Traumatology Arthroscopy, Clinical Orthopaedics and Related Research* for the orthopaedic field), grey literature of unpublished trials and a manual search or the reference list of the papers included in the meta-analysis. As clinical trials with positive results have more chance of being published, there could be a publication bias. For this reason, many journals now require every published RCT to be registered at the clinicaltrials.gov website before their execution, to promote the tracking of each RCT despite its positive or negative results (Fact Box 47.3). This source should also be searched to minimise the possibility of missing relevant results.

In terms of the search strategy, the choice of the key terms should aim to produce the most extensive search possible. However, it is necessary to strike a balance between striving for

> **Fact Box 47.3: What clinicaltrials.gov is?**
>
> ClinicalTrials.gov was created because of the Food and Drug Administration Modernization Act of 1997. It is a Web-based resource that provides patients, their family members, healthcare professionals, researchers and the public with easy access to information on publicly and privately supported clinical studies on a wide range of diseases and conditions. Each ClinicalTrials.gov record presents summary information about a study protocol including disease, intervention, study design and contact information for the study locations and, for some records, also includes information on the results of the study.

comprehensiveness and maintaining relevance, as increasing the comprehensiveness of a search will reduce its precision and retrieve more non-relevant articles. Usually, three sets of terms, developed for the healthcare condition, intervention(s) and study design, should be combined using the Boolean "AND" operator. To be as comprehensive as possible, it is necessary to include a wide range of free-text terms for each of the selected concepts (e.g. "injury", "rupture" or "lesion" related to the ACL or Achilles tendon), combined using the Boolean "OR" operator. The function of "truncation" (e.g. Menisc* for Meniscus or Meniscal) could be a strategy for maximising the results. As this is one of the most important phases of the development of a meta-analysis, the help of a librarian specialising in database searches could be useful.

47.4 How to Obtain the Appropriate Data

47.4.1 Define Precise Inclusion and Exclusion Criteria

The most common reference manager software enables the pooling together of all the results obtained from the systematic search in the

selected databases and the discarding of all the duplicates. To prepare a precise flow chart of study selection, all the numbers of studies found, removed and excluded should be noted and justified.

The crucial step in the phase of study selection is a clear and precise definition of inclusion and exclusion criteria [26]. Their goal is to create a homogeneous study population for the meta-analysis. The rationale for their choice should be stated, as it may not be apparent to the reader. Inclusion criteria may be based on study design, sample size and characteristics of the subject, type of treatment and follow-up. Examples of exclusion criteria include studies not published in English or as full-length manuscripts, drop-out rate (usually >20%), doctoral theses and studies published in non-peer-reviewed journals. Moreover, the outcomes that should be analysed and the way they are presented could be a matter of inclusion and exclusion criteria (e.g. radiographic evaluation of osteoarthritis according only to the Kellgren-Lawrence scale). This search and data extraction (presented below) should be performed by two independent investigators; the results should therefore be compared and any disagreement should be resolved by a third independent investigator. Usually, the first screening of all the results is based on title and abstract evaluation. The full text of the potentially eligible studies should then be obtained and carefully evaluated.

47.4.2 Extract All the Relevant Information

Two or more authors of a meta-analysis should abstract information from studies independently. "Data" is defined as any information about (or deriving from) a study, including details of methods, participants, setting, context, interventions, outcomes, results, publications and investigators [13]. Data should only be collected from separate sets of patients, and the authors should be careful to avoid studies that publish the same subjects or overlapping groups of subjects that appeared in different studies in duplicate publications. In this

Table 47.1 Summary of the main items that should be extracted from each study included in a meta-analysis

Main items considered in data collection for a meta-analysis	
Methods	Study design and duration, sequence generation, allocation sequence concealment, blinding, other concerns regarding bias, inclusion and exclusion criteria
Participants	Total number, setting, diagnostic criteria, age, gender, co-morbidities
Interventions	Number of intervention groups, specific interventions, intervention details
Outcomes	Outcomes and time points, definition, unit, scale
Results	Number of participants in each intervention group and for each outcome, summary data for each intervention group, eventual subgroup analysis
Miscellaneous	Funding source, key conclusion from the study authors

case, the wider population or most recent report can be chosen. The main items considered in data collection are presented in Table 47.1:

If all the relevant information cannot be obtained from the full text, the study authors could be contacted to request the missing desired information. This step could be crucial if several parameters are missing in the outcome report, since these data are fundamental for the statistical calculation of the meta-analysis. For dichotomous outcomes, only the number of patients that experience the outcome and the total number of patients are needed. Moreover, categorical values (e.g. Kellgren-Lawrence scale for knee osteoarthritis) should be analysed as dichotomous variables, after defining a clinically meaningful cut-off value to determine two groups. On the other hand, for continuous data, the mean values and especially the SDs are needed as well. Since the SD is mandatory for effect-size calculation, if it is not reported in the original study, there are several artefacts to approximate it from the known information, such as standard error (SE), the range of values or the p-value [14] (Table 47.2). The last option is to impute the SD, borrowing the SD from a similar study included in the meta-analysis,

Table 47.2 Methods to derive the standard deviation from the data usually provided in a scientific paper

Obtain standard deviation from the available data
Obtaining standard deviation from the standard error
Standard deviation: $\text{Standard error} \times \sqrt{\text{sample size}}$
Obtaining standard deviation from the ranges
Standard deviation: (Upper range − lower range)/4
Obtaining standard deviation from the *p*-value
Step 1: from p-value to t-value
Excel formula: =tinv(*P*-value,degrees of freedom)
Degrees of freedom isThe number of patients in an experimental group + control group − 2
Step 2: from t-value to standard error
Standard error: (mean of group 1 − means of group 2)/*t*-value
Step 3: from standard error to standard deviation
Standard deviation: $\text{Standard error} / \sqrt{(1/\text{patient of group one} + 1/\text{group two})}$

using the highest or a "reasonably high" SD, or using the average SD. All imputation techniques involve making assumptions about unknown statistics introducing a source of bias, and it is best to avoid using them wherever possible. If most of the studies in a meta-analysis have missing standard deviations, these values should not be imputed, and the meta-analysis cannot be performed. A narrative presentation of the results is then preferable. On the other hand, when SDs are derived, a sensitivity analysis (see further section) is recommended to establish whether the derived imprecision could affect the result of the meta-analysis and the effect of a treatment.

47.5　How to Analyse the Obtained Data

47.5.1　Choose Appropriate Statistical Software

The tabulation of data and the calculation of simple medians, means and SDs are possible with Microsoft Excel or the equivalent. However, to perform an appropriate statistical analysis of a meta-analysis, dedicated software is necessary. One of the most frequently used is Review Manager (RevMan, Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014), the official software of the Cochrane Community, which is available free of charge and which has been designed to facilitate preparation of protocols and full reviews, including text, characteristics of studies, comparison tables and study data. Moreover, it can perform meta-analysis of the data entered and present the results graphically. Another valid free alternative is OpenMetaAnalyst, which is a simple open-source software with an Excel-like interface for performing meta-analyses of binary, continuous or diagnostic data, using a variety fixed- and random-effect methods, including Bayesian and maximum likelihood analysis. It also enables to perform meta-regression analysis, meta-analysis of proportions and continuous variables from single-arm studies and cumulative, leave-one-out or subgroup analyses. Another simple software is MedCalc (MedCalc Software, Ostend, Belgium), which offers also the possibility to perform a multitude of statistical tests and analysis, with a variety of graphic representations. However, its options for meta-analysis purposes are limited, and moreover it is only available freely as a sample.

47.5.2　Define Correct Effect-Size Measurement

As previously stated, the outcomes could be presented in two ways: dichotomous or continuous. Based on this, the effect size of the treatment

should be presented properly, using one or more of the following parameters [5, 17].

The risk ratio (RR): also defined as the relative risk, is the ratio of the risk of an event in the two groups. It is used for dichotomous outcomes and ranges from 0 to infinity. The RR describes the multiplication of the risk that occurs using the experimental intervention. For example, an RR of 5 for a treatment implies that events with treatment are five times more likely than events with the control treatment. Alternatively, we can say that the experimental treatment increases the risk of events by $100 \times (RR - 1)\% = 400\%$. Conversely, an RR of 0.20 is interpreted as the probability of an event with experimental treatment being a fifth of that with control treatment. Alternatively, this reduction could be expressed using the relative risk reduction (RRR) (see below). The interpretation of the clinical importance of a given RR should be made considering the typical risk of events with the control treatment, since an RR of 5 could correspond to a clinically important increase in events from 10 to 50%, or a small, less clinically important increase from 1 to 5%.

The relative risk reduction (RRR): is an alternative way of re-expressing the RR as a percentage of the reduction of the relative risk after the experimental treatment compared with the control treatment. For example, an RR of 0.20 is interpreted as the probability of an event with experimental treatment being a fifth of that with control treatment. Since the RRR is calculated as $100 \times (1 - RR)\%$, we can therefore say that the experimental treatment reduces the risk of events by 80%.

The odds ratio (OR): considering "odds" as the ratio between the probability that a particular event will occur and the probability that it will not occur, the OR is the ratio of the odds of an event in the two groups. It is used for dichotomous outcomes and ranges from 1 to infinity. The OR is more difficult to interpret, as it describes the multiplication of the odds of the outcome that occur using the experimental intervention. To understand what an OR means in terms of changes in the numbers of events, it is simplest to convert to it into an RR and then interpret the risk ratio in the context of a typical control group risk. Attention should be paid to not misinterpreting RR and OR.

The risk difference (RD): is defined as the difference between the observed risks (proportions of individuals with the outcome of interest) between the two groups of experimental and control treatment. It is used for dichotomous outcomes and could be comprised between -1 and $+1$ (which means that it could be easily converted to per cent by multiplying it by 100). Like the RR, the clinical importance of an RD may depend on the underlying risk of events in the control treatment, since an RD of 0.05 (or 5%) may represent a small, clinically insignificant change from a risk of 55 to 60% or a proportionally much larger and potentially important change from 1 to 6%.

The number needed to treat (NNT): is defined as the expected number of people who need to receive the experimental treatment to obtain one additional person either incurring or avoiding the considered event. It is used for dichotomous outcomes and is always a positive number, usually rounded up to the nearest whole number. It is derived from the absolute value of |RD| and is calculated as 1/|RD|. An RD of 0.23 therefore corresponds to an NNT of 4.34, which is rounded up to 5, and it means that "it is expected that one additional (or less) patient will incur the event for every five participants receiving the experimental intervention rather than control". Since the NNT gives an "expected value", it does not imply that one additional event will necessarily occur in every group of "*n*" patients treated with the experimental procedure.

The mean difference (MD): is defined as the absolute difference between the mean value of the experimental and control group. It is used for continuous outcome measured with the same scale and estimates the amount by which the experimental intervention changes the outcome on average compared with the control.

The standardised mean difference (SMD): is defined as the size of the intervention effect in each study relative to the variability observed in the study. Since it is used when the same outcome is measured with different scales (e.g. knee function

measured with Lysholm, subjective IKDC or KOOS scores), the results should be statistically standardised to a uniform scale before being combined. Care should be taken when the scales have a different "direction" (e.g. a scale increase while the others decrease with disease severity); in this case, multiplying for −1 should be used when needed, to ensure that all the scales point in same direction.

47.5.3 Identify and Measure Heterogeneity

Heterogeneity is a term used to describe the variability of studies. Variability in the studied participants (e.g. males or females, old adult or adolescent patients), interventions (open or minimally invasive procedures, different grafts for reconstructions) and outcomes (objective or subjective) may be described as clinical heterogeneity, variability in study design and risk of bias may be described as methodological heterogeneity, while the variability in the treatment effects in the different studies is known as statistical heterogeneity [17, 21]. The last one is usually a consequence of clinical or methodological diversity, or both, between the studies. Studies with methodological flaws and small studies may overestimate treatment effects and can contribute to statistical heterogeneity. The statistical heterogeneity should therefore be examined and quantified using statistical tests, to implement measures to reduce the risk of bias [4]. The chi-square test (χ^2) assesses whether observed differences in results are compatible with chance alone: a low p-value provides evidence of heterogeneity in intervention effects. Since this measurement did not provide the "amount" of heterogeneity, quantification according to the Higgins I^2 statistic should be performed. This test produces a 0–100% value that represents the percentage of total variation across studies due to heterogeneity. According to the Cochrane Guidelines, this value is interpreted as follows: 0–40% heterogeneity might be not important, 30–60% may represent moderate heterogeneity, 50–90% may represent substantial heterogeneity and 75–100% considerable heterogeneity. However, there are

no empirically developed cut-off points to determine when there is too much heterogeneity to perform a meta-analysis, and it is left to the author's discretion to determine whether a meta-analysis is appropriate, based on the results of the test of heterogeneity and clinical judgement.

47.5.4 Apply Strategies to Address High Heterogeneity: The Random Effect and Others
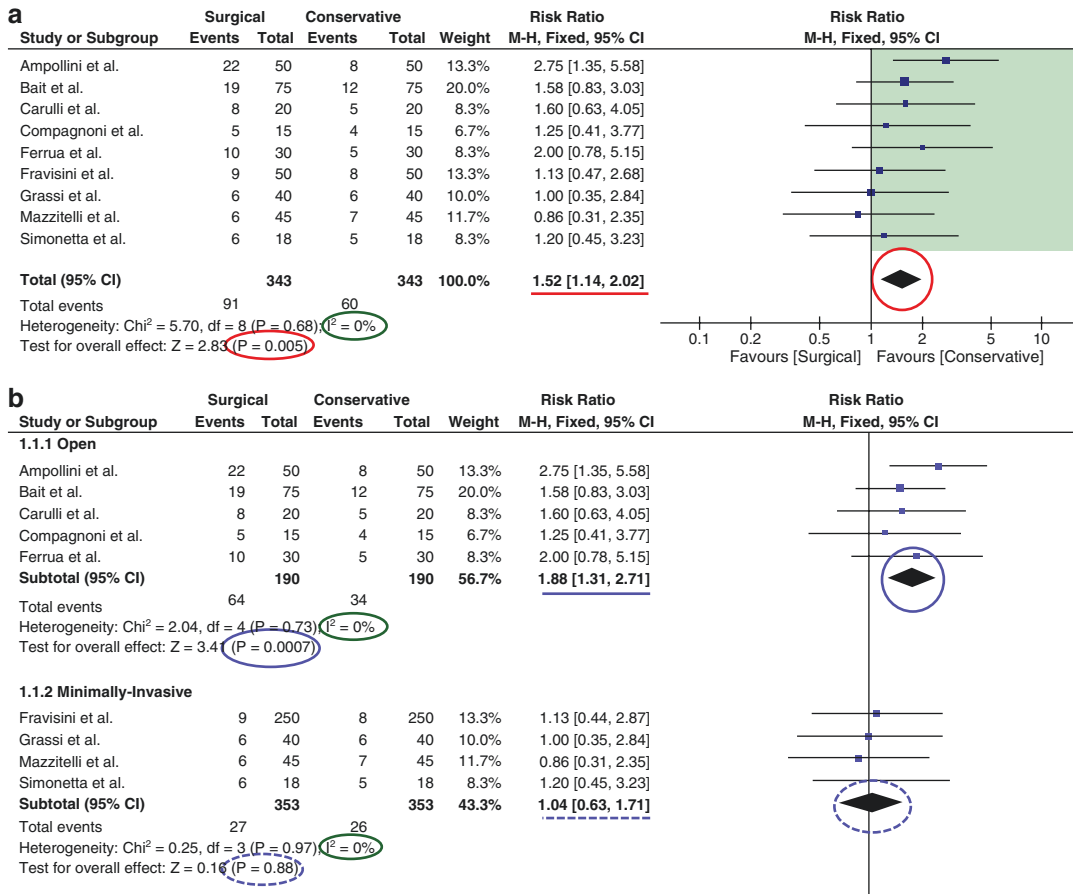
Since high heterogeneity implies dissimilarity in the studies, a meta-analysis should be conducted with caution, and several strategies should be implemented to consider this situation. These should be applied after checking whether data are correct and no errors have been made in data extraction [4, 5].

Perform random-effect meta-analysis: the two most frequently used models to conduct a meta-analysis are the fixed- (Mantel-Haenszel, inverse variance or Peto methods) [24] and random-effect (DerSimonian and Laird method) [6] models. The fixed-effect model investigates the question "What is the best estimate of the population effect size?", assuming a common treatment effect and that the differences between studies are due to chance. This model should be used with low heterogeneity, as it gives greater weight to larger studies. On the other hand, the randomeffect model investigates the question "What is the average treatment effect?", assuming the distribution of the treatment effect along a range of values. This model should be used with high heterogeneity that cannot be explained, as it assumes that the treatment effect in the different studies is not identical due to clinical and methodological heterogeneity, and this model therefore gives less weight to larger studies. If the heterogeneity is not extreme, they frequently lead to similar results. On the other hand, they can produce different conclusions, and the use of fixed- or random-effect models should therefore be carefully considered based on the amount of heterogeneity, despite no guidelines existing in this direction (random effect is usually used with $I^2 > 50\%$) (Fig. 47.1).
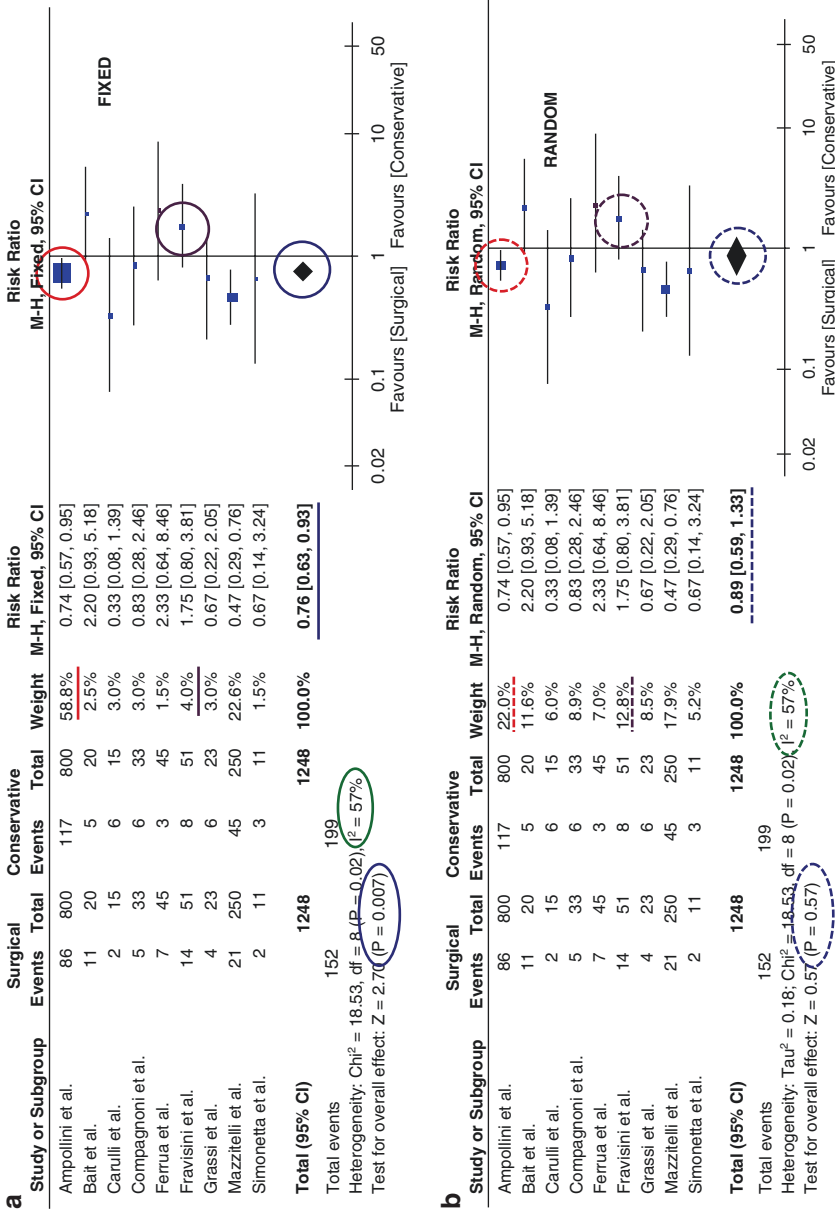
Fig. 47.1 Example of forest plots evaluating the risk ratio of treatment failure between the conservative and surgical treatment for a fictitious pathology. If a fixed-effect method was used, larger studies are represented with the largest squares (red circle) proportional to their weight (red line); similarly, the smallest studies have small squares (purple circle) and weight (purple line). In this specific case, the diamond of the overall effect (blue circle) crosses the central line and its confidence intervals do not contains the null value (RR = 1); it could therefore be assumed that there is a significant reduction in failure risk after surgical treatment (P.0.007). However, the results should be interpreted with extreme caution due to the high and significant heterogeneity (green circle) (**a**). If a random-effect method is used to account for the high heterogeneity, the final result changes dramatically: the largest studies and smallest studies are given less weight (red dotted line and circle) and more weight (purple dotted line and circle), respectively, and this produces an enlargement of confidence intervals for the overall effect, which crosses the central line containing the null value (RR = 1). So, when random effect was used, we can affirm that there is no evidence of a significant effect of surgical treatment compared with conservative treatment in reducing failures (P = 0.57) (**b**)

a

Study or Subgroup	Surgical Events	Total	Conservative Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Ampollini et al.	22	50	8	50	13.3%	2.75 [1.35, 5.58]
Bait et al.	19	75	12	75	20.0%	1.58 [0.83, 3.03]
Carulli et al.	8	20	5	20	8.3%	1.60 [0.63, 4.05]
Compagnoni et al.	5	15	4	15	6.7%	1.25 [0.41, 3.77]
Ferrua et al.	10	30	5	30	8.3%	2.00 [0.78, 5.15]
Fravisini et al.	9	50	8	50	13.3%	1.13 [0.47, 2.68]
Grassi et al.	6	40	6	40	10.0%	1.00 [0.35, 2.84]
Mazzitelli et al.	6	45	7	45	11.7%	0.86 [0.31, 2.35]
Simonetta et al.	6	18	5	18	8.3%	1.20 [0.45, 3.23]
Total (95% CI)		**343**		**343**	**100.0%**	**1.52 [1.14, 2.02]**
Total events	91		60			

Heterogeneity: Chi^2 = 5.70, df = 8 (P = 0.68); I^2 = 0%
Test for overall effect: Z = 2.83 (P = 0.005)

Risk Ratio M-H, Fixed, 95% CI — 0.1 0.2 0.5 1 2 5 10 — Favours [Surgical] Favours [Conservative]

b

Study or Subgroup	Surgical Events	Total	Conservative Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
1.1.1 Open						
Ampollini et al.	22	50	8	50	13.3%	2.75 [1.35, 5.58]
Bait et al.	19	75	12	75	20.0%	1.58 [0.83, 3.03]
Carulli et al.	8	20	5	20	8.3%	1.60 [0.63, 4.05]
Compagnoni et al.	5	15	4	15	6.7%	1.25 [0.41, 3.77]
Ferrua et al.	10	30	5	30	8.3%	2.00 [0.78, 5.15]
Subtotal (95% CI)		**190**		**190**	**56.7%**	**1.88 [1.31, 2.71]**
Total events	64		34			

Heterogeneity: Chi^2 = 2.04, df = 4 (P = 0.73); I^2 = 0%
Test for overall effect: Z = 3.41 (P = 0.0007)

	Surgical Events	Total	Conservative Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI
1.1.2 Minimally-Invasive						
Fravisini et al.	9	250	8	250	13.3%	1.13 [0.44, 2.87]
Grassi et al.	6	40	6	40	10.0%	1.00 [0.35, 2.84]
Mazzitelli et al.	6	45	7	45	11.7%	0.86 [0.31, 2.35]
Simonetta et al.	6	18	5	18	8.3%	1.20 [0.45, 3.23]
Subtotal (95% CI)		**353**		**353**	**43.3%**	**1.04 [0.63, 1.71]**
Total events	27		26			

Heterogeneity: Chi^2 = 0.25, df = 3 (P = 0.97); I^2 = 0%
Test for overall effect: Z = 0.16 (P = 0.88)

Fig. 47.2 Example of forest plot evaluating the risk ratio of complications after the conservative or surgical treatment for a fictitious pathology (**a**). In this case, it appears to be correct to use a fixed-effect model since the heterogeneity is null (green circle); this is confirmed by the concordance of the effect size of most of the studies, which lies on the right side of the forest plot. Since the confidence intervals do not contain the null value (RR = 1) (red circle and line), it could be suggested that there is evidence of an increased risk of complications after surgical treatment for the fictitious pathology. However, if a sub- group analysis is performed separating open and minimally invasive surgery, the results differ from the main analysis (**b**). In the case of open surgery, the final overall effect is similar to the main analysis (blue line and circle); in the case of minimally invasive surgery, the confidence intervals of the overall effect (dotted blue line and circle) contain the null value (RR = 1), suggesting that there is no evidence of a difference in complications after minimally invasive surgery or conservative treatment for the fictitious pathology

Perform a subgroup analysis: subgroup analyses may be conducted as a means of investigating heterogeneous results or to answer specific questions about patient groups, types of intervention or types of study (Fig. 47.2). They can be performed for a subset of participants or a subset of studies, and they consider the meta-analysis results from each group separately. The non-overlap of the CI usually indicates statistical significance and a different effect of the treatment within the subgroups, thereby explaining the heterogeneity to some extent. However, the subgroup analysis should be limited only to restricted cases, since it could increase the risk of type-II error due to the reduction of patient cohort size.

Perform a meta-regression: meta-regression is an extension of subgroup analyses that allows the investigation of the effect of continuous or categorical characteristics simultaneously. Its

role is like that of simple regression, where the outcome variable is the effect estimate (RR, OR, MD), and the explanatory variables, or covariates, are the study characteristics that might influence the size of the intervention effect. The regression coefficient obtained from a meta-regression will describe how the outcome variable changes with a unit increase in the covariate, while its statistical significance describes whether there is a linear relationship between them. It should be underlined that the characteristics to investigate (covariates) should be justified by biological and clinical hypotheses and should be the lowest number possible. One limitation of the meta-regression is that more than ten studies in the meta-analysis are generally required for its use.
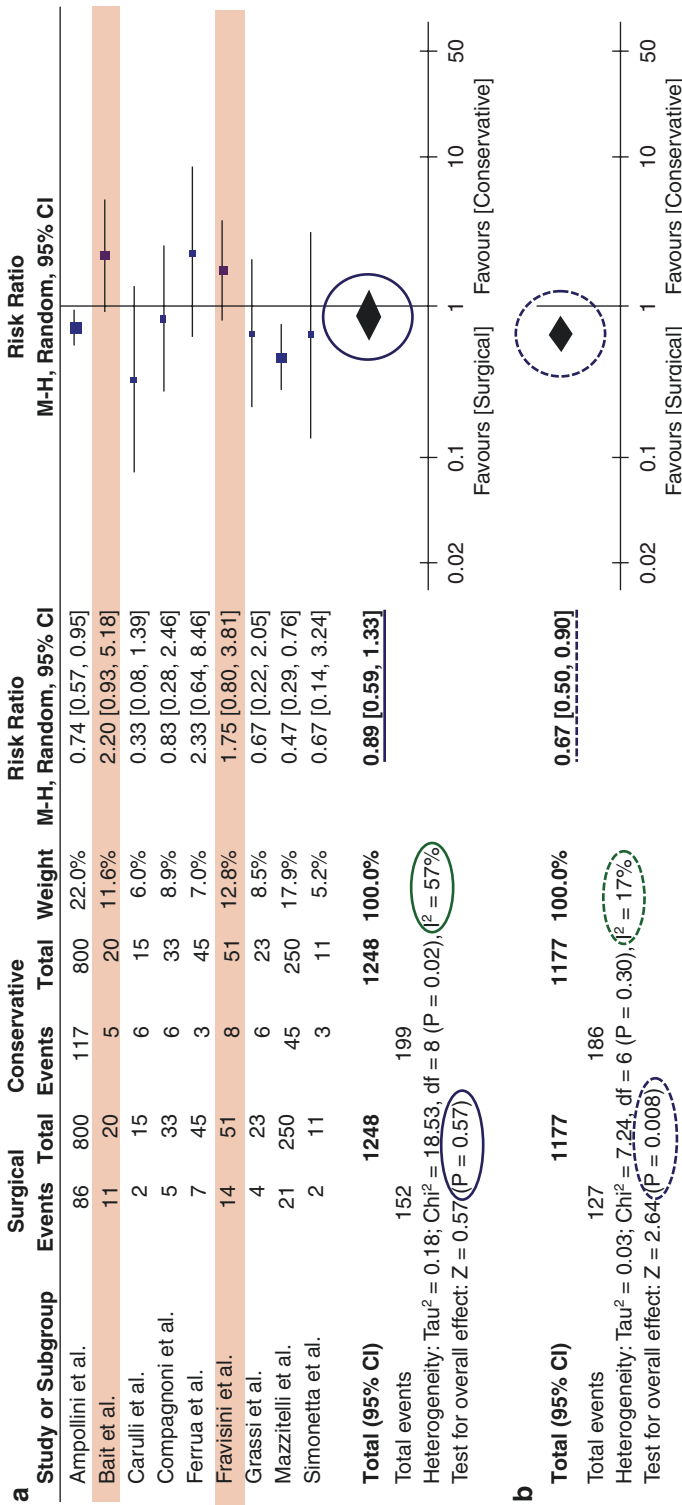
Perform a sensitivity analysis: while the aim of the subgroup analysis is to estimate a treatment effect for a particular subgroup, the aim of the sensitivity analysis is to investigate whether the meta-analysis findings change based on different arbitrary or unclear decisions related to the meta-analysis process. The main decisions that can generate the need for a sensitivity analysis could be related to the eligibility criteria of the studies (e.g. study design or methodological issues), the data analysed (e.g. imputation of missing SD) or analysis methods (fixed or random effect, choice of effect-size measurement). In practical terms, the sensitivity analysis consists of the repetition of the meta-analysis, excluding the studies burdened by unclear or arbitrary decisions and in the informal comparison of the different ways the same thing is estimated. After the sensitivity analysis, when the overall conclusions are not affected by the different decisions made during the review process, more certainty can be assumed. On the other hand, if decisions are identified as influencing the findings, the results must be interpreted with an appropriate degree of caution if it is not possible to improve the process (Fig. 47.3). As different from the

subgroup analysis, the sensitivity analysis is not designed to estimate the effect of the intervention in the group excluded from the analysis, so its report should be produced with a summary table.

Change the measurement of effect: the choice of the measurements of effect size may affect the degree of heterogeneity; however, it is unclear whether the heterogeneity of intervention effect alone is a suitable criterion for choosing between the different measurements.

Exclude studies: since heterogeneity could be due to the presence of one or two outliers, studies with conflicting results compared with the rest could be excluded, as their exclusion could address the problem of heterogeneity. However, is not appropriate to exclude a study based on its result since it may introduce a bias. It can therefore only be removed with confidence if there are obvious reasons. Unfortunately, there are no tests to determine the extent of clinical heterogeneity, and researchers must decide whether the studies contributing to a meta-analysis are similar enough clinically to make meta-analysis feasible. Refining inclusion criteria and excluding studies, even if this reduces heterogeneity, also decreases the total number of articles included on a topic. A sensitivity analysis is suggested to check whether the excluded study/studies could alter the meta-analysis results.

Do not perform a meta-analysis: if high heterogeneity cannot be addressed using the presented strategies, the investigator should consider whether the amount of heterogeneity is so large that the results of the meta-analysis are problematic. In this case, especially when there is inconsistency in the direction of the treatment effects that could make the use of an average value misleading, meta-analysis should be abandoned, and the evidence should be fairly expressed in a systematic review. Another and frequent reason to avoid meta-analytic pooling of data is when too few studies with no new findings are obtained after the systematic search.

a

Study or Subgroup	Surgical Events	Surgical Total	Conservative Events	Conservative Total	Weight	Risk Ratio M-H, Random, 95% CI
Ampollini et al.	86	800	117	800	22.0%	0.74 [0.57, 0.95]
Bait et al.	11	20	5	20	11.6%	2.20 [0.93, 5.18]
Carulli et al.	2	15	6	15	6.0%	0.33 [0.08, 1.39]
Compagnoni et al.	5	33	6	33	8.9%	0.83 [0.28, 2.46]
Ferrua et al.	7	45	3	45	7.0%	2.33 [0.64, 8.46]
Fravisini et al.	14	51	8	51	12.8%	1.75 [0.80, 3.81]
Grassi et al.	4	23	6	23	8.5%	0.67 [0.22, 2.05]
Mazzitelli et al.	21	250	45	250	17.9%	0.47 [0.29, 0.76]
Simonetta et al.	2	11	3	11	5.2%	0.67 [0.14, 3.24]
Total (95% CI)		**1248**		**1248**	**100.0%**	**0.89 [0.59, 1.33]**
Total events	152		199			

Heterogeneity: Tau² = 0.18; Chi² = 18.53, df = 8 (P = 0.02); I² = 57%
Test for overall effect: Z = 0.57 (P = 0.57)

b

						Risk Ratio M-H, Random, 95% CI
Total (95% CI)		**1177**		**1177**	**100.0%**	**0.67 [0.50, 0.90]**
Total events	127		186			

Heterogeneity: Tau² = 0.03; Chi² = 7.24, df = 6 (P = 0.30); I² = 17%
Test for overall effect: Z = 2.64 (P = 0.008)

Fig. 47.3 Example of forest plots evaluating the risk ratio of treatment failure between conservative and surgical treatment for a fictitious pathology (**a**). After a random-effect meta-analysis due to the high heterogeneity (green circle), the final result is that there is no evidence of effect of surgical treatment in reducing the risk of failures, since the confidence intervals of the overall effect contain the null value (RR = 1). However, after risk-of-bias assessment, two studies with poor methodology and a high risk of bias have been found (Bait et al. and Fravisini et al.) and excluded through a sensitivity analysis (**b**). The results of the sensitivity analysis show a reduction in heterogeneity (green dotted circle) and the narrowing of the confidence intervals of the overall effect (blue dotted line and circle) that no longer contain the null value (RR = 1). In this case, the results of the meta-analysis should be interpreted with extreme caution since methodological issues and biases were able to affect the entity of the overall effect of treatment

47.6   How to Present and Evaluate the Results

47.6.1   Prepare the Forest-Plot Graphic for the Main Outcomes

The result section of a meta-analysis should summarise the findings in a clear, logical order, explicitly addressing the objective of the review [11, 30]. The characteristics of methods, participants, intervention and outcomes should be reported in a narrative manner or with reference to tables [31]. On the other hand, the data analysis is better presented through the so-called "forest-plot" graphic; this is a simple, immediate and visually friendly method for describing the raw data, estimate and CI of the chosen effect measurement, the choice between fixed- or random-effect meta-analysis, the heterogeneity, the weight of each study and a test for the overall effect (Fig. 47.1). Forest plots should not be used when an outcome has only been investigated in a single study.

The measurement of the effect of each study included in the meta-analysis and in the forest plot is represented by a square, with the dimension proportional to its weight (based on sample size and the choice of a fixed- or random-effect model) and a horizontal line corresponding to its CI. Since the CI describes a range of values within which we can be reasonably sure that the true effect lies, a narrow CI indicates an effect size that is known precisely, while a very wide CI indicates that we

have little knowledge of the effect. The CI width for an individual study depends on the sample size, SD (for continuous outcomes) and risk of the event (dichotomous outcomes). When the CI crosses the central line (indicating an MD or SMD of 0 and an OR or RR of 1), it is possible that the experimental or control treatment has the same effect on the evaluated outcome. If the effect size of most of the included studies lies on the same side of the graphic, thus indicating a similar effect of the treatment, the overall heterogeneity is usually low.

The overall effect size of the meta-analysis is represented by a "diamond". Its position indicates the value of the effect size, while its width indicates the CI. This width depends on the precision of the individual study estimates, the number of studies combined and the heterogeneity (in random-effect models, precision will decline with increasing heterogeneity). When the 95% CI for the effect of the meta-analysis does not cross the central line, it excludes the null value (MD or SMD of 0 and OR or RR of 1), and the *p*-value of the overall meta-analysis will therefore be <0.05. In this case, we can affirm that the observed effect is very unlikely to have arisen purely by chance and, as a result, there are differences in the effect of experimental and control interventions.

The forest plot, in a certain study design, could present the effect size of continuous or dichotomous outcomes from single-arm case series; in this case, we only have the estimation of pooled outcomes, without the comparison between two treatments (Fig. 47.4).

Studies	Estimate (95% C.I.)		
Ampollini et al.	90 . 000	(87 . 189,	92 . 811)
Baitet al.	86 . 000	(84 . 131,	87 . 869)
Carulli et al.	88 . 000	(85 . 370,	90 . 630)
Compagnoni et al.	85 . 000	(77 . 301,	92 . 699)
Ferrua et al.	89 . 000	(86 . 093,	91 . 907)
Fravisini et al.	93 . 000	(90 . 977,	95 . 023)
Grassi et al.	91 . 000	(88 . 470,	93 . 530)
Mazzitelli et al.	90 . 000	(88 . 891,	91 . 109)
Simonetta et al.	95 . 000	(85 . 382,	104 . 618)
Overall (I^2=74.24 % ,P< 0.001)	**89 . 547**	**(87 . 872,**	**91 . 222)**
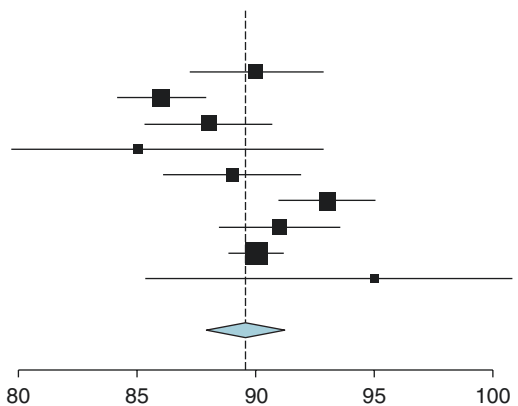
Fig. 47.4 Example of a forest plot of a continuous outcome from single-arm case series. In this case, no comparison between treatments is made, since a single treatment is evaluated in the included studies; the main result is therefore the mean of the considered score or outcome, based on the weight of each study

47.6.2 Perform a Methodological Assessment and Bias Evaluation

The fundamental measurement to guarantee credibility in a meta-analysis is the evaluation of the methodology and bias of the included studies: this is a necessary step that should not be missed, because it could generate a misinterpretation of the results [12]. First, the level of evidence (Table 47.3) should be immediately and clearly reported. For single-arm case series, many methodological questionnaires are available, and one of the most frequently used is the Coleman Score [2] or its modifications [22] (Table 47.4). For non-randomised controlled studies, authors usually refer to the Newcastle-Ottawa Scale (NOS) [34] or its modifications (Table 47.5). For RCTs, there is also a vast choice of scores and checklists [27], with those obtained from the Consolidated Standards for Reporting Trials (CONSORT) guidelines regarded as some of the most authoritative [16] (Table 47.6). However, the indispensable action to ensure scientific strictness is the risk-of-bias evaluation, which is performed using the "Cochrane Risk of Bias Tool" [12].

A bias is defined as a systematic error (or a deviation from the truth) in results, and it can lead to an underestimation or overestimation of the true intervention effect. The types of bias considered in the Cochrane Risk of Bias Tool are:

- Selection bias: the systematic difference between the baseline characteristics of the groups

Table 47.3 List of the five levels of evidence

Level of evidence for therapeutic studies	
Level I	Randomised controlled trial (RCT)
Level II	Prospective cohort studies (non-randomised comparative study)
Level III	Retrospective cohort study (non-randomised comparative study); case-control study
Level IV	Case series
Level V	Mechanism-based reasoning

- Performance bias: the systematic difference between groups in the care that is provided
- Attrition bias: the systematic difference between groups in withdrawals from a study
- Detection bias: the systematic difference between groups in how outcomes are determined
- Reporting bias: the systematic difference between reported and unreported findings

According to these types of bias, seven domains are evaluated and rated as a low, unclear and high risk of bias.

- *Random sequence generation (selection bias)*: describe the method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it should produce comparable groups.
- *Allocation concealment (selection bias)*: describe the method used to conceal the allocation sequence in sufficient detail to determine whether intervention allocations could have been foreseen in advance of, or during, enrolment.
- *Blinding of participants and personnel (performance bias)*: describe all the measures used, if any, to blind study participants and personnel to knowledge of the intervention a participant received. Provide any information relating to whether the intended blinding was effective. An evaluation should be made for each main outcome.
- *Blinding of outcome assessment (detection bias)*: describe all the measures used, if any, to blind outcome assessors to knowledge of the intervention a participant received. Provide any information relating to whether the intended blinding was effective. An evaluation should be made for each main outcome.
- *Incomplete outcome data (attrition bias)*: describe the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis. State whether attrition and exclusions were reported, the numbers in each intervention group (compared with total randomised participants), reasons for attrition/exclusions where reported

Table 47.4 Modified Coleman Score used for orthopaedic case series (adapted from: Magnussen RA, Carey JL, Spindler KP. Does autograft choice determine intermediate-term outcome of ACL reconstruction? Knee Surg Sports Traumatol Arthrosc. 2011;19(3):462–472)

Modified Coleman Methodology Score		
Outcome	Option	Points
Part A: Only one score to be given for each section (total = 60)		
1. Study size: number of patients	>120	10
	81–120	7
	40–80	4
	<40 or not stated	0
2. Mean follow-up (years)	>6 years	5
	3–6 years	3
	<3 years, not stated, unclear	0
3. Percentage of patients with follow-up	>90%	5
	80–90%	3
	<80%	0
4. No. of interventions per group or separate outcomes should be reported	One procedure	10
	More than one procedure but consistent	5
	Among all patients in each group	
	Unclear or multiple interventions	0
	Among patients in the same group	
5. Type of study	Randomised controlled trial	15
	Prospective cohort study	10
	Retrospective cohort study	5
6. Diagnostic certainty	In all	5
	In >80%	3
	In <80%, not stated, unclear	0
7. Description of surgical technique	Technique stated with details	5
	Technique named without elaboration	3
	Not stated, unclear	0
8. Description of postoperative rehabilitation	Well described	5
	Described without complete detail	3
	Protocol not reported	0
Part B: Scores could be given for each option in each of the three sections (total = 40)		
1. Outcome criteria	Outcome measurements clearly defined	2
	Timing of outcome assessment clear	2
	Use of outcome with good reliability	3
	Use of outcome with good sensitivity	3
2. Procedure for assessing outcomes	Subjects recruited	5
	Independent investigators	4
	Written assessment	3
	Patient-centred data collected	3
3. Description of subject selection process	Selection criteria reported and unbiased	5
	Recruitment rate reported >80%	5
	Eligible subjects not included in the	5
	Study satisfactorily accounted for	

The total score (max = 100) of the evaluated study is calculated as the sum of Parts A and B

Table 47.5 Modified Newcastle-Ottawa Scale for non-randomised studies (adapted from: http://www.uphs.upenn.edu/cep/methods/Modified%20Newcastle-Ottawa.pdf)

Question	Y\N
Study population:	
1. All study groups derived from similar source/reference populations?	
2. Attrition not significantly different across study groups?	
Study validity:	
3. The measurement of exposure is valid?	
4. The measurement of outcome is valid?	
5. Investigators blinded to end-point assessment?	
Confounders:	
6. Potential confounders identified (e.g. co-morbidities)	
7. Statistical adjustment for potential confounders made?	
8. Funding source(s) disclosed and no obvious conflict of interests?	
The scale does not require a calculation of a numeric score but can be used to map the study characteristics visually	

Table 47.6 Reporting quality scale for randomised controlled trials based on the Consolidated Standards for Reporting Trials (CONSORT) guidelines (adapted from: Huwiler-Müntener K, Jüni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. JAMA. 2002 Jun 5;287(21):2801–4)

Reporting quality scale based on 1996 CONSORT statement	
Question	Y\N
1. Does the title identify the study as a randomised controlled trial?	
2. Is the abstract presented in structured format?	
3. Are objectives stated?	
4. Is hypothesis stated?	
5. Is the study population described?	
6. Are the inclusion and exclusion criteria described?	
7. Are the interventions described?	
8. Are the outcome measurements described?	
9. Is a primary outcome specified?	
10. Is a minimum (clinically?) important difference for the primary outcome reported?	
11. Are power calculations described?	
12. Is the rationale for the statistical analyses explained?	
13. Are the methods for statistical analyses described?	
14. Are stopping rules described?	
15. Is the unit of randomisation described?	
16. Is the method used to generate the allocation schedule described?	
17. Is the method of allocation concealment described?	
18. Is the timing of assignment described?	
19. Is the method for separating those generating the allocation sequence from those assigning participants to groups described?	
20. Are the mechanisms of blinding described?	
21. Is the number of eligible patients reported?	
22. Is the number of randomised patients reported for each comparison group?	
23. Are prognostic variables by treatment and control group described?	
24. Is the number of patients receiving an intervention as allocated reported for each group?	
25. Is the number of patients analysed reported for each comparison group?	
26. Are withdrawals and drop-outs described for each comparison group?	
27. Are protocol deviations described for each comparison group?	
28. Is the estimated effect of intervention on primary and secondary outcomes stated, including measurements of precision?	
29. Are the results stated in absolute numbers?	
30. Are summary data and inferential statistics presented in sufficient detail to permit alternative analyses and replication?	
The total score (max = 30) of the evaluated study is calculated as the sum of the answer "Yes"	

Table 47.7 Strategies to identify high, low or unclear risk of bias for the main domains of the "Cochrane Risk of Bias Tool" (adapted from: Table 8.5c of Higgins JPT, Altman DG (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons, 2008)

Risk-of-bias assessment according to the "Cochrane Risk of Bias Tool"		
Selection	*Random sequence generation:* method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it should produce comparable groups	
	Low risk	Randomisation with random number table, coin toss, computer generator, dice…
	High risk	Randomisation with date of birth, admission day, clinic record number, judgement of clinician…
	Unclear risk	Insufficient information on randomisation process
	Allocation concealment: method used to conceal the allocation sequence to avoid intervention allocations being foreseen in advance or during enrolment	
	Low risk	Participants could not foresee assignation due to central allocation, identical drug containers, opaque sealed envelopes
	High risk	Participants could foresee assignation due to open allocation schedule, alternation, unsealed envelopes
	Unclear risk	Insufficient information or concealment not described
Performance	*Blinding of participants and personnel:* measures used to blind study participants and personnel from knowing the intervention received	
	Low risk	Blinding or use of outcomes not influenced by lack of blinding
	High risk	No or incomplete blinding that could influence outcomes
	Unclear risk	Insufficient information to permit judgement
Detection	*Blinding of outcome assessment:* measures used to blind outcome assessors from knowledge of the intervention a participant received	
	Low risk	Blinding or use of outcomes not influenced by lack of blinding
	High risk	No or incomplete blinding that could influence outcomes
	Unclear risk	Insufficient information to permit judgement
Attrition	*Incomplete outcome data:* completeness of outcome data for each main outcome, including attrition and exclusion from the analysis	
	Low risk	No missing outcome data, or missing data balanced between groups and not related to outcomes
	High risk	Reason for missing data related to outcomes, or imbalance between intervention group
	Unclear risk	Insufficient reporting of exclusions to permit judgement
Reporting	*Selective reporting:* possibility of selective outcome reporting	
	Low risk	All the prespecified outcomes of interest have been reported
	High risk	Not all prespecified outcomes have been reported, or outcomes reported incompletely, or lack of key outcomes
	Unclear risk	Insufficient information to permit judgement
Other	*Other bias:* any important concerns about bias not addressed in the other domains in the tool	
	Low risk	The study appears to be free from other sources of bias
	High risk	Bias related to specific study design, or fraudulent, or other problems
	Unclear risk	Insufficient information to assess whether other biases exist

and any re-inclusions in analyses performed by the review authors.

- *Selective reporting (reporting bias):* state how the possibility of selective outcome reporting was examined by the review authors and what was found.

- *Other (other bias):* state any important concerns about bias not addressed in the other domains in the tool.

According to the quality and methodology of the study, each domain should be rated (Table 47.7). The overall risk of bias should

Fact Box 47.4: Bias Included and Evaluated in the "Cochrane Risk of Bias Tool"

- **Selection bias**: the systematic difference between the baseline characteristics of the groups
- **Performance bias**: the systematic difference between groups in the care that is provided
- **Attrition bias**: the systematic difference between groups in withdrawals from a study
- **Detection bias**: the systematic difference between groups in how outcomes are determined
- **Reporting bias:** the systematic difference between reported and unreported findings

therefore be determined based on a low risk of bias for all key domains (low risk), a high risk of bias for one or more key domain (high risk) or an unclear risk for one or more key domain (unclear risk). Finally, the risk across studies could be defined as low if most information comes from studies with a low risk of bias, high if the proportion of information from studies with a high risk of bias is sufficient to affect the interpretation of the results and unclear if most information is from studies with a low or unclear risk of bias. To help the presentation of this information, a risk-of-bias summary (Fig. 47.5) and risk-of-bias graphs (Fig. 47.6) are extremely useful.

Other types of bias exist, due to imbalance in the dissemination of research findings due to the nature and direction of results. They are known as reporting biases [35] and can be:

- Publication bias: when the publication or non-publication of research findings depends on the nature and direction of the results; as an example, studies with negative results are often not published
- Time-lag bias: when the rapid or delayed publication of research findings depends on the nature and direction of the results

|	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Ampollini et al.	+	+		+	−	+	
Bait et al.	+	+	+	+	−		+
Carulli et al.	−	−	−	−	+	+	−
Compagnoni et al.			+	+	+	+	−
Ferrua et al.	+	−	+	+		+	
Fravisini et al.	+	+	−	−	+	+	
Grassi et al.	+	+			+	+	+
Mazzitelli et al.			−	−	+	+	
Simonetta et al.	+		+		+		−

Fig. 47.5 In this risk-of-bias summary, it is possible to have a visual presentation of the risk of each bias for all the included studies. The risk could be low (green plus), high (red minus), or unclear (no sign). In this specific table, it is possible to observe that the study by Carulli et al. is the most biased and the study by Grassi et al. is the one with the lowest risk of bias

- Multiple publication bias (duplicate): when the multiple or singular publication of research findings depends on the nature and direction of the results
- Location bias: when the publication of research findings in journals with different ease of access or levels of indexing in standard databases depends on the nature and direction of results
- Citation bias: when the citation or non-citation of research findings depends on the nature and direction of the results
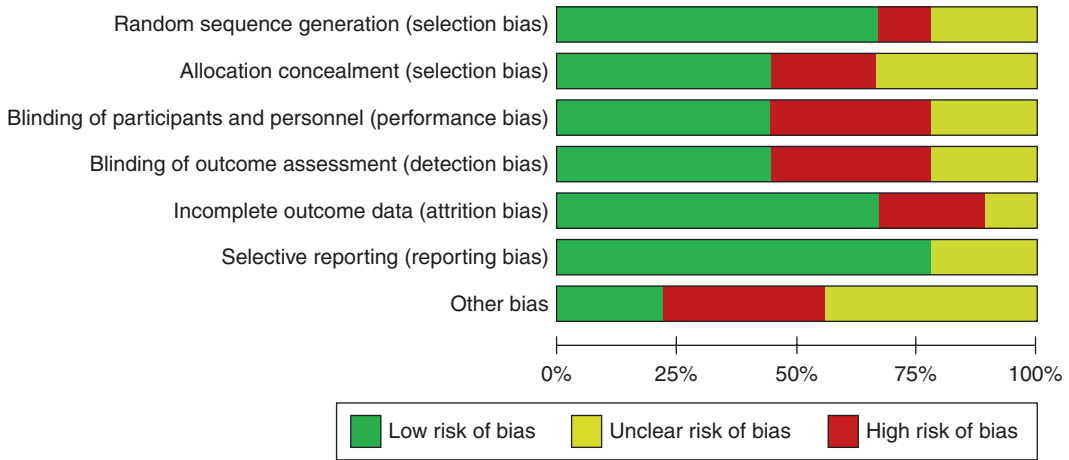
Fig. 47.6 In these risk-of-bias graphs, it is possible to see a visual presentation of the recurrent bias based on each domain. In this specific case, the "reporting bias" appears to be the bias with a lower risk, while the "performance bias" and the "detection bias" appear to be those with a higher risk
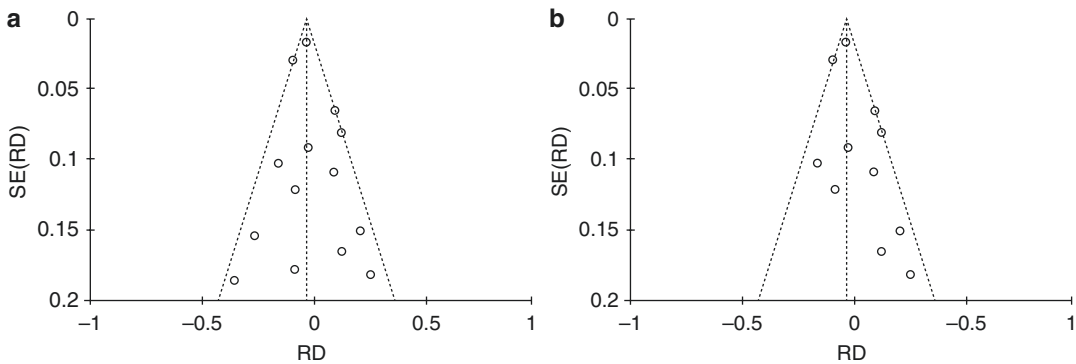
Fig. 47.7 In a funnel plot where the risk of bias is low, the symmetrical shape of an inverted funnel is seen (**a**). When the publication bias tends to exclude the publication of small studies without statistical significance, the funnel plot results are asymmetrical, with a gap in the bottom corner (bottom left corner, in this case) (**b**)

- Language bias: when the publication of research findings in a language other than English are sometimes be regarded as of secondary importance, while studies publishing positive results might also be more likely to publish in English

Outcome reporting bias: when the selective reporting of some outcomes but not others depends on the type of results found, if they are positive or negative or if they introduce new or repetitive findings.

One practical way to detect reporting bias is the use of the funnel plot graph [35]. This is a simple scatter plot of the intervention effect estimates from individual studies against some measurement of each study's size of precision; the effect estimates are plotted on the horizontal scale and the measurement of study size on the vertical axis (Fig. 47.7). As effect estimates from small studies scatter more widely at the bottom of the graph and those of larger studies are scattered more narrowly at the top of the graph, the plot should assume the shape of a symmetrical inverted funnel. If, due to publication bias, smaller studies without statistical significance remain unpublished, the plot will be asymmetrical with a gap in a bottom corner. In this case, the

meta-analysis will tend to overestimate the intervention effect. Apart from publication bias, asymmetry of the funnel plot could also be due to poor methodological quality, true heterogeneity, artefactual or chance. Funnel plot asymmetry should only be used if at least ten studies are included in the meta-analysis, when all the studies do not have similar sizes.

47.6.3 Correctly Approach and Evaluate Non-randomised Studies

Finally, a few words should be devoted to the meta-analysis of non-randomised controlled studies. Pooling together the results of non-randomised studies could be appropriate when they have a large effect; however, combining RCT and non-randomised studies is not recommended, as their results should be expected to differ systematically, resulting in increased heterogeneity [29].

Meta-analyses of non-randomised studies have greater potential bias, and their results should therefore be interpreted with caution. In fact, serious concerns could be related especially to the differences between people in different intervention groups (selection bias), caused by the lack of randomisation.

If both RCTs and non-randomised studies of an intervention are available and the author also wants to include a non-randomised study due to the small number of RCTs, they should be presented separately, or the findings of the non-randomised studies should be discussed in the final discussion with the meta-analysis findings.

47.7 How to Interpret Your Findings Critically

47.7.1 Summarise Your Main Findings

After the results have been correctly and clearly reported and methodology and bias adequately evaluated, the main findings of the meta-analysis can be critically interpreted.

For this purpose, "summary of findings" tables could be useful, since it presents the main findings in a simple format, providing key information on the quality of evidence, the magnitude of the effect of the interventions and the sum of available data on the main outcomes [31]. Six elements should be reported: a list of all important outcomes, a measurement of the typical burden of these outcomes, the absolute and relative magnitude of effect, numbers of participants and studies addressing these outcomes, a rating of the overall quality of evidence for each outcome and a space for comments. Special mention should be made of the quality of evidence, which is assessed through the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) tool [1]. It describes the body of evidence as "High", "Moderate", "Low" or "Very Low", based on the methodological quality, directness of evidence, heterogeneity, precision of effect estimates and risk of publication bias.

47.7.2 Pay Attention: What Is "Statistically Significant" Is Also "Clinically Significant"

When numerical results are going to be interpreted, attention should be paid to the 95% CI, because, if it is narrow, the effect size is known precisely, while, if it is wider, the uncertainty is greater. The CI and the *p*-value of the meta-analysis are strictly linked, as a value of <0.05 will exclude the null value (OR, RR of 1 or MD, SMD of 0) from the interval between the CIs, thus suggesting that the experimental treatment has an effect compared with the control treatment.

However, even if the findings are statistically significant, the clinical meaning of the benefit of the experimental treatment should be accurately weighted. When the treatment effect is measured with RR or RD, an interpretation of the clinical importance cannot be made without knowledge of the typical risk of events without treatment. In fact, a risk ratio of 0.75, for example, could correspond to a clinically important reduction in events from 80 to 60% or a small, less clinically important reduction from 4 to 3%. Conversely,

when dealing with continuous scales and mean differences, the proper minimum clinically important difference (MCID) of the considered outcome should be considered. Since the MCID represents the smallest change in a treatment outcome that a patient would identify as important, it is possible that a mean difference, despite being statistically significant, could be irrelevant from a clinical point of view (e.g. an MD of 4 points in the subjective IKDC, where the MCID is 11.5 points) [3, 8] (Table 47.8). Moreover, the minimum detectable change (MDC), which is the minimum amount of change in a patient's score that ensures the change is not the result of measurement error, should be considered [3]. Recently, in a JBJS commentary, the superiority of double-bundle ACL reconstruction compared with single-bundle demonstrated in a Level I RCT has been questioned, since a difference of less than 1 mm in an arthrometric evaluation and around two points in a subjective IKDC were not considered clinically meaningful, despite being statistically significant [23].

Furthermore, conclusions should not be drawn too quickly without performing an accurate evaluation of heterogeneity through subgroup or sensitivity analysis. For example, Soroceanu et al. [33] reported a relative risk of re-rupture of 0.4 in favour of surgical repair compared with conservative treatment in the case of Achilles tendon rupture. However, since the authors found a not negligible heterogeneity of 35%, they identified the item of "functional rehabilitation" as a cause of heterogeneity through meta-regression. So,

Table 47.8 The minimum detectable change (MDC) and minimum clinically important difference (MCID) of the main clinical scores used in knee surgery (adapted from: Collins NJ1, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS) and Tegner Activity Score (TAS). Arthritis Care Res (Hoboken). 2011 Nov;63 Suppl 11:S208–28. doi: https://doi.org/10.1002/acr.20632)

The clinical scores most used for knee evaluation			
Score	Condition	MDC	MCID
Subjective IKDC	Injuries	8.8–15.6	6.3 (6 m) to 16.7 (12 m)
Subjective IKDC	Mixed pathologies	6.7	11.5 (sensitive) to 20.5 (specific)
KOOS pain	Injuries	6–6.1	–
KOOS symptoms	Injuries	5–8.5	–
KOOS ADL	Injuries	7–8	–
KOOS sport/rec	Injuries	5.8–12	–
KOOS Qol	Injuries	7–7.2	–
KOOS pain	OA	13.4	–
KOOS symptoms	OA	15.5	–
KOOS ADL	OA	15.4	–
KOOS sport/rec	OA	19.6	–
KOOS Qol	OA	21.1	–
Lysholm	Injuries	8.9–10.1	–
Lysholm	Mixed pathologies	–	–
Oxford Knee Scale	OA	6.1	–
WOMAC pain	OA	14.4–16.2	22.87 (TKR 6 m) to 27.98 (TKR 24 m)
WOMAC symptoms	OA	22.9–30.6	14.43 (TKR 6 m) to 21.35 (TKR 24 m)
WOMAC function	OA	10.6–15	19.01 (TKR 6 m) to 20.84 (TKR 24 m)
Tegner Activity Scale	Injuries	1	–
Tegner Activity Scale	OA	–	–

MDC minimum detectable change, *MCID* minimum clinically important difference

after performing a subgroup analysis separating patients undergoing functional or conventional rehabilitation, they found no difference in the re-rupture rate between surgical treatment and conservative treatment with functional rehabilitation. On the other hand, as Foster et al. [7] intended to include as many data as possible in their meta-analysis of irradiated vs. nonirradiated allografts for ACL reconstruction, they performed a sensitivity analysis to evaluate whether imputing SD would have influenced the final results. After performing an analysis of only the studies reporting SD, they repeated the analysis also adding those studies in which the SD was imputed as the mean value, reporting no substantial differences in the results. In their final evaluation, they therefore disclosed this issue and presented the data relative to all the studies, independently of the source of the SD.

47.7.3 Translate Your Findings into Clinics

The final difficulty in interpreting the meta-analysis results lies in applying the results to clinical practice [32]. It is important to correctly disclose whether the individual studies pooled in the meta-analysis can be generalised to a specific clinical scenario. This includes ensuring similar patient populations, interventions and outcomes of interest. For example, Jiang et al. [18] reported no differences in the rates of return to sport between patients undergoing surgical repair or non-surgical treatment after Achilles tendon rupture. However, since the RCTs included in the meta-analysis evaluated patients with a mean age of around 40 years, this recommendation should be applied with extreme caution in young, professional athletes.

Finally, attention should be paid when interpreting inconclusive or counter-intuitive results, which is one of the most common errors in scientific manuscripts. When there is inconclusive evidence, it is not appropriate to state that "there is evidence of no effect". It is instead more appropriate to state that "there is no evidence of an effect".

When results are instead counter-intuitive, clinical judgement based on experience, education and current practices will be needed to decipher the unexpected results. The decision to determine whether to accept the findings or question the statistical technique could be taken after looking back at the original articles, reassessing their inclusion and evaluating whether assumptions about the original research question are not lost when the studies are combined.

47.8 Conclusion: How to Prepare the Manuscript

The very last step in meta-analysis is to prepare a manuscript that is complete, essential, clear to the reader and suitable for publication in a peer-reviewed journal. First, the guideline of the target journal should be consulted to "tailor" the manuscript accordingly. Then, the PRISMA guidelines (Table 47.9) should be followed to fulfil the highest quality standard [25].

Title: should be concise and focused on the topic, identifying the paper as a meta-analysis.

Abstract: should be structured, including all the sections of the paper.

Introduction: should be short and focused on the topic, expressing the rationale of the meta-analysis, the purpose and the hypothesis.

Methods: should mention all the information regarding the databases used, the timing of the search, the keywords, the eligibility criteria, the methods of data extraction and the items evaluated. The statistical method used to combine the results, the bias evaluation and eventual sensitivity or subgroup analysis should be mentioned as well.

Results: should be clear and easy to understand. All included and excluded studies should be described in a flow diagram. It is recommended to present the data through forest plots and summary tables. The results of sensitivity or subgroup analysis and of bias evaluation should be provided in this section.

Discussion: should not be too long and should preferably focus on the main findings of the meta-analysis. The evidence should be sum-

Table 47.9 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist as guideline for the final production of a meta-analysis (adapted from: Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009 Jul 21;339:b2700)

Preferred reporting items for systematic reviews and meta-analyses (PRISMA) checklist			
Section	Item	Description	Page
Title	1	Title	Identify the report as a systematic review, meta-analysis, or both
Abstract	2	Structured summary	Include, as applicable: background; objectives; data sources; eligibility criteria, participants and interventions; synthesis methods; results; limitations; conclusions
Introduction	3	Rationale	Describe the rationale for the review in the context of what is already known
	4	Objectives	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes and study design (PICOS)
Methods	5	Protocol and registration	Indicate if a review protocol exists and, if available, provide registration information including registration number
	6	Eligibility criteria	Specify study characteristics and report characteristics used as criteria for eligibility, giving rationale
	7	Information sources	Describe all information sources and date last searched
	8	Search	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated
	9	Study selection	State the process for selecting studies
	10	Data collection process	Describe the method of data extraction from reports
	11	Data items	List and define all variables for the data that were sought and any assumptions and simplifications made
	12	Risk of bias in individual studies	Describe the methods used for assessing risk of bias of individual studies and how this information is to be used in any data synthesis
	13	Summary measurements	State the principal summary measurements (e.g. risk ratio, difference in means)
	14	Synthesis of results	Describe the methods for handling data and combining results of studies
	15	Risk of bias across studies	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g. publication bias)
	16	Additional analyses	Describe methods of additional analyses (e.g. sensitivity or subgroup analyses, meta-regression)
Results	17	Study selection	Give numbers of studies screened, assessed for eligibility and included in the review, with reasons for exclusion at each stage (flow diagram)
	18	Study characteristics	For each study, present characteristics of the data that were extracted
	19	Risk of bias within studies	Present data on risk of bias of each study
	20	Results of individual studies	For all outcomes considered, present for each study simple summary data for each intervention group and effect estimates with CI (forest plot)
	21	Synthesis of results	Present results of each meta-analysis conducted, including confidence intervals and measurements of consistency.
	22	Risk of bias across studies	Present results of any assessment of risk of bias across studies
	23	Additional analysis	Give results of additional analyses, if performed (e.g. sensitivity or subgroup analyses, meta-regression)

Table 47.9 (continued)

Preferred reporting items for systematic reviews and meta-analyses (PRISMA) checklist				
Section	Item	Description		Page
Discussion	24	Summary of evidence	Summarise the main findings including the strength of evidence for each main outcome	
	25	Limitations	Discuss limitations at study and outcome level and at review level	
	26	Conclusions	Provide a general interpretation of the results in the context of other evidence and implications for future research	
Funding	27	Funding	Describe sources of funding for the systematic review and other support; role of funders for the systematic review	

marised and discussed, together with the main limitations. The conclusions, which must be based exclusively on the findings without speculation, should delineate clinical and research implications.

Figures and tables: forest plots and funnel plots are very useful for result presentation, such as summary tables.

References: should be updated and formatted according to journal guidelines.

Funding: authors should always disclose any conflict of interests and eventual funding.

Clinical Vignette

After attending an International conference, you discover that the attention of Sports Medicine surgeons is on a new device to treat a specific type of ankle fractures, which has been introduced few years ago. You are aware of a couple of pilot RCT and, after performing a quick PubMed search, you find out at least three new RCTs published in the last year and a completed trial on clinicaltrials.gov, which is held by some of your overseas colleagues. Therefore, you plan to perform a systematic search to run a meta-analysis comparing this new device with the standard of care. With the help of your librarian and an orthopaedic resident in your hospital, you can define a broad and appropriate search strategy, analysing three databases and the website clinicaltrials.gov. Defining as inclusion criteria only RCT comparing the new device with the classic approach, you can find nine studies which

are pooled in a formal meta-analysis with the help of the biostatistician of your university. Due to the several differences in surgical procedure and patients' inclusion criteria, you opted to perform a more conservative statistical analysis using a random-effect model, considering also the high degree of statistical heterogeneity revealed with the I^2 test. Analysing the relative risk of complication and the mean differences of the main disease-specific scales, you find out a significant superiority of the new device. However, when you performed the bias evaluation, the lack of blinding of patients and clinicians raises some concerns due to the high risk of detection and performance bias. Overall, after the critical evaluation of results and bias, you agree with the chief of your clinic that the implementation of this new device in your clinical practice, with specific indications, could improve the quality of your treatments.

Take-Home Message

- Meta-analysis can be a powerful tool to combine results from studies with similar design and patient populations that are too small or underpowered individually.
- However, there are many potential threats that can limit the internal validity and real clinical impact of conclusions reported in a meta-analysis.
- An appropriate study question and design, the proper management of heterogeneity and the methodological evaluation of included studies with bias assessment are necessary to ensure the highest quality.

Appendix: Internet Links and Websites Useful for the Various Steps in Preparing a Meta-Analysis

Appendix: useful links	
Guides to meta-analyses	
Cochrane Handbook for Systematic Review of Interventions	http://handbook.cochrane.org/
PRISMA Guidelines for Systematic Reviews and Meta-Analyses	http://www.prisma-statement.org/
GRADE Handbook	http://gdt.guidelinedevelopment.org/app/handbook/handbook.html
Databases	
Cochrane Library	http://www.thecochranelibrary.com
PubMed	https://www.ncbi.nlm.nih.gov/pubmed/
Embase	http://store.elsevier.com/embase
Clinical Trials Database	https://clinicaltrials.gov/
Statistical software	
Cochrane RevMan	http://tech.cochrane.org/revman/download
OpenMetaAnalyst	http://www.cebm.brown.edu/openmeta/download.html
MedCalc	https://www.medcalc.org/download.php
Methodological evaluation	
Oxford Level of Evidence	http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/
JBJS Level of Evidence	http://jbjs.org/level-of-evidence
Newcastle-Ottawa Scale for Non-randomised Studies	http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
Modified Newcastle-Ottawa Scale	http://www.uphs.upenn.edu/cep/methods/Modified%20Newcastle-Ottawa.pdf
CONSORT checklist for randomised controlled trials (RCT)	http://www.consort-statement.org/
PEDRO scale for randomised controlled trials (RCT)	https://www.pedro.org.au/english/downloads/pedro-scale/
AMSTAR score for systematic reviews	https://amstar.ca/Amstar_Checklist.php
COSMIN guidelines for studies of measurement instrument	http://www.cosmin.nl/downloads.html (for studies of measurement instruments)

References

1. Brozek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, Phillips B, Lelgemann M, Lethaby A, Bousquet J, Guyatt GH, Schünemann HJ, GRADE Working Group. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. Allergy. 2009;64(5):669–77 . Review. https://doi.org/10.1111/j.1398-9995.2009.01973.x.

2. Coleman BD, Khan KM, Maffulli N, Cook JL, Wark JD. Studies of surgical outcome after patellar tendinopathy: clinical significance of methodological deficiencies and guidelines for future studies. Victorian Institute of Sport Tendon Study Group. Scand J Med Sci Sports. 2000;10:2–11.

3. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). Arthritis Care Res (Hoboken). 2011;63(Suppl 11):S208–28. https://doi.org/10.1002/acr.20632.

4. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. 2nd ed. London: BMJ Publication Group; 2001.

5. Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008.

6. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7:177–88.

7. Foster TE, Wolfe BL, Ryan S, Silvestri L, Kaye EK. Does the graft source really matter in the outcome of patients undergoing anterior cruciate ligament reconstruction? An evaluation of autograft versus allograft reconstruction results: a systematic review. Am J Sports Med. 2010;38(1):189–99. https://doi.org/10.1177/0363546509356530.

8. Grassi A, Ardern CL, Marcheggiani Muccioli GM, Neri MP, Marcacci M, Zaffagnini S. Does revision ACL reconstruction measure up to primary surgery? A meta-analysis comparing patient-reported and clinician-reported outcomes, and radiographic results. Br J Sports Med. 2016;50(12):716–24. https://doi.org/10.1136/bjsports-2015-094948.

9. Grassi A, Zaffagnini S, Marcheggiani Muccioli GM, Neri MP, Della Villa S, Marcacci M. After revision anterior cruciate ligament reconstruction, who returns to sport? A systematic review and meta-analysis. Br J Sports Med. 2015;49(20):1295–304. https://doi.org/10.1136/bjsports-2014-094089.

10. Grassi A, Zaffagnini S, Marcheggiani Muccioli GM, Roberti Di Sarsina T, Urrizola Barrientos F, Marcacci M. Revision anterior cruciate ligament reconstruction does not prevent progression in one out of five patients of osteoarthritis: a meta-analysis of prevalence and progression of osteoarthritis. J ISAKOS. 2016;1(1):16–24. https://doi.org/10.1136/jisakos-2015-000029.

11. Greco T, Zangrillo A, Biondi-Zoccai G, Landoni G. Meta-analysis: pitfalls and hints. Heart Lung Vessel. 2013;5(4):219–25. Review.

12. Higgins JPT, Altman DG. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008.

13. Higgins JPT, Deeks JJ. Chapter 7: Selecting studies and collecting data. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008.

14. Higgins JPT, Deeks JJ, Altman DG. Chapter 16: Special topics in statistics. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008.

15. Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008.

16. Huwiler-Müntener K, Jüni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. JAMA. 2002;287(21):2801–4.

17. Israel H, Richter RR. A guide to understanding meta-analysis. J Orthop Sports Phys Ther. 2011;41(7):496–504. https://doi.org/10.2519/jospt.2011.3333.

18. Jiang N, Wang B, Chen A, Dong F, Yu B. Operative versus nonoperative treatment for acute Achilles tendon rupture: a meta-analysis based on current evidence. Int Orthop. 2012;36(4):765–73. https://doi.org/10.1007/s00264-011-1431-3.

19. Koretz RL, Lipman TO. Understanding systematic reviews and meta-analyses. JPEN J Parenter Enteral Nutr. 2016. pii: 0148607116661841.

20. Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008.

21. Lefaivre KA, Slobogean GP. Understanding systematic reviews and meta-analyses in orthopaedics. J Am Acad Orthop Surg. 2013;21(4):245–55 . Review. https://doi.org/10.5435/JAAOS-21-04-245.

22. Magnussen RA, Carey JL, Spindler KP. Does autograft choice determine intermediate-term outcome of ACL reconstruction? Knee Surg Sports Traumatol Arthrosc. 2011;19(3):462–72.

23. Marx RG. Anatomic double-bundle anterior cruciate ligament reconstruction was superior to conventional single-bundle reconstruction. J Bone Joint Surg Am. 2013;95(4):365. https://doi.org/10.2106/JBJS.9504.ebo804.

24. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst. 1959;22:719–48.

25. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535. https://doi.org/10.1136/bmj.b2535. No abstract available.

26. O'Connor D, Green S, Higgins JPT. Chapter 5: Defining the review question and developing criteria for including studies. In: Higgins JPT, Green S, editors. Cochrane handbook of systematic reviews of interventions. Chichester: Wiley; 2008.

27. Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther. 2008;88(2):156–75.

28. Parlamas G, Hannon CP, Murawski CD, Smyth NA, Ma Y, Kerkhoffs GM, van Dijk CN, Karlsson J, Kennedy JG. Treatment of chronic syndesmotic injury: a systematic review and meta-analysis. Knee Surg Sports Traumatol Arthrosc. 2013;21(8):1931–9. Review. https://doi.org/10.1007/s00167-013-2515-y.

29. Reeves BC, Deeks JJ, Higgins JPT, Wells GA. Chapter 13: Including non-randomized studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008.

30. Russo MW. How to review a meta-analysis. Gastroenterol Hepatol (N Y). 2007;3(8):637–42.
31. Schunemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Chapter 11: Presenting results and 'summary of findings' tables. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008.
32. Schunemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, Guyatt GH. Chapter 12: Interpreting results and drawing conclusions. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008.
33. Soroceanu A, Sidhwa F, Aarabi S, Kaufman A, Glazebrook M. Surgical versus nonsurgical treatment of acute Achilles tendon rupture: a meta-analysis of randomized trials. J Bone Joint Surg Am. 2012;94(23):2136–43. https://doi.org/10.2106/JBJS.K.00917.
34. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol. 2010;25(9):603–5. https://doi.org/10.1007/s10654-010-9491-z.
35. Sterne JAC, Egger M, Moher D. Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008.
36. Wright JG, Swiontkowski MF, Tolo VT. Meta-analyses and systematic reviews: new guidelines for JBJS. J Bone Joint Surg Am. 2012;94(17):1537. No abstract available.