

The Value of Systematic Reviews and Meta-Analyses in Surgery

Simon R. Knight

Centre for Evidence in Transplantation, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK

Keywords

Systematic review · Meta-analysis · Surgery

Abstract

Background: Systematic reviews and meta-analyses are generally regarded as sitting atop the hierarchy of clinical evidence. The unbiased summary of current evidence that a systematic review provides, along with the increased statistical power from larger numbers of patients, is invaluable in guiding clinical decision-making and development of practice guidelines. Surgical specialties have historically lagged behind other areas of medicine in the application of evidence-based medicine, perhaps due to the unique challenges faced in the conduct of surgical clinical trials. These challenges extend to the conduct of systematic reviews, due to issues with the quality and heterogeneity of the underlying literature. **Summary:** Recent years have seen an improvement in the quality of randomized controlled trials in surgical topics and an explosion in the publication of systematic reviews. This review explores recent trends in systematic reviews in surgery and discussed some of the aspects in conducting and interpreting reviews that are unique to surgical topics, including blinding, surgical heterogeneity and learning curves, patient and clinician preference, and industry involvement. **Key Messages:** Clinical trials, and therefore sys-

tematic reviews, of surgical interventions pose unique challenges which are important to consider when conducting them or applying the findings to clinical practice. Despite the challenges, systematic reviews still represent the best level of evidence for development of surgical practice guidelines.

© 2021 S. Karger AG, Basel

Introduction

The central principle of evidence-based medicine (EBM) is the unbiased application of best current evidence to guide the treatment of patients [1]. Alongside robust clinical trials, the systematic review is a key component of EBM and sits atop the hierarchy of evidence described by the Oxford Centre for Evidence-Based Medicine [2]. As with other medical specialties, systematic reviews play a key role in guiding clinical practice and informing guidelines.

Whilst many of the principles for systematic review and meta-analysis are common across all areas of medicine, clinical trials (and therefore systematic reviews) for surgical interventions have specific nuances that can add additional challenges. These include aspects such as learning curves, between-surgeon heterogeneity, challenges with blinding, and evaluation of emergency inter-

ventions. Perhaps for this reason, uptake of EBM in surgical specialties has typically lagged behind other areas of medicine, and even now it is widely accepted that the quality of both clinical trials and systematic reviews in surgery is variable at best [3–5]. This review has discussed the key components and benefits of systematic reviews, focussing specifically on the challenges faced by researchers in surgical disciplines.

What Is a Systematic Review?

A systematic review is a structured and explicit attempt to identify, critically appraise, and synthesize all of the available evidence to answer a specific clinical question. The standardized and transparent methodology described for systematic reviews aims to reduce the risk of bias in study selection, data extraction, and synthesis. This contrasts with an “expert review,” where there is the potential to introduce the reviewer’s own biases by unrepresentative selection of evidence or biased interpretation.

“Meta-analysis” refers to the statistical combination of results from multiple studies to provide a summary effect. Meta-analysis forms only a small part of a systematic review, and indeed is not a requirement. Where data are heterogeneous or sparse; a narrative synthesis of the evidence can be preferable to formal meta-analysis, whilst still adhering to the rest of the principles of systematic review.

A properly conducted systematic review is a piece of scientific research and should follow a pre-defined protocol. To reduce the risk of reviewer bias, it is recommended that the review protocol is registered in advance in a public repository such as the Prospective Register of Systematic Reviews (PROSPERO, <https://www.crd.york.ac.uk/prospéro/>) in much the same way that clinical trial protocols are prospectively registered. This helps to ensure that the protocol is not changed as the reviewers identify the evidence.

Again analogous to clinical trials, there are well-defined reporting guidelines that ensure that the strict methodology required for an unbiased review is adhered to. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines have been adopted by the International Committee of Medical Journal Editors and the checklist should be completed by all review authors prior to submission [6].

When conducted well, a systematic review can be very resource intensive and time-consuming, requiring mul-

iple reviewers and often months or years of time. However, the results are often far more powerful than any individual clinical trial and provide the most robust evidence that we have for guiding clinical practice.

The Added Value of a Systematic Review

Often, multiple clinical trials, reportedly of the same intervention, provide conflicting evidence as to the benefits and risks. By identifying all of the available evidence, the systematic review process allows the reviewer to explore the differences between study populations and interventions and try to explain these conflicting findings. By incorporating greater numbers of patients, statistical power is increased, allowing exploration of subgroups not possible in individual studies and sometimes providing more information on secondary endpoints that the original studies had power to identify. If the evidence is found to be consistent across a number of clinical trials in different healthcare settings, this adds confidence in the generalizability of findings over individual studies.

A good example of the value of a systematic review can be seen in the case of mechanical bowel preparation prior to colorectal surgery. The use of bowel preparation was historically common practice, with the belief that it would reduce the risk of anastomotic leak and infectious complications. An early clinical trial in 1972 questioned this practice, and subsequent studies suggested that it may even be harmful [7]. The majority of trials in this area are small, with very few recruiting above 100 patients, meaning that the ability to demonstrate equivalence in outcomes without bowel preparation is limited. The effect of bowel preparation on anastomotic leak rates varies dramatically between trials, with low event rates and odds ratios ranging from 0.37 to 8.54. A Cochrane review on the topic synthesized all of this evidence, and found that in 18 trials including 5,805 patients, there was no difference in the risk of leak, death, reoperation, or infectious complications, supporting the view that preparation is not required for colonic surgery [8]. Current clinical practice guidelines suggest that mechanical bowel preparation should be avoided for colonic surgery [9].

The results of systematic reviews are not always so equivocal. Often, a paucity of evidence or poor quality in the underlying evidence means that firm conclusions cannot be drawn as to the benefits of a given intervention. This does not, however, mean that the review is not of value. By identifying the existing evidence and understanding the deficiencies in what is already known, such

Table 1. The stages of a systematic review

Stage	Description
Defining the question	Clear definition of the review question, usually in PICO format Population: considering patient age, underlying disease/stage, and co-morbidities Intervention: considering drug or procedure, dose, concomitant care, and duration Comparator: no treatment, placebo, and current gold standard Outcomes: consider efficacy, safety, and health economics Study design: RCT or observation studies
Searching the literature	Comprehensive literature search across multiple databases, using keywords and free text terms. Searches of trial databases and reference lists
Quality assessment	Assessment for risk of bias, including selection bias, performance bias, measurement bias, attrition bias, reporting bias, and publication bias
Data extraction and synthesis	Extraction of structured data for all outcomes. Narrative synthesis or meta-analysis depending on quality of underlying evidence Assessment for heterogeneity, with attempts to explain where identified
Summary and conclusion	Interpretation of findings in light of clinical and statistical significance and heterogeneity. Application to real-world practice

PICO, Population, Intervention, Comparator, Outcome; RCTs, randomized controlled trials.

a review can be invaluable in securing funding and directing the design of future clinical trials to add to the evidence-base.

The Steps of a Systematic Review

A robust systematic review usually follows 5 main steps (Table 1).

Step 1: Defining the Question

Albert Einstein is often quoted as having said “If I had an hour to solve a problem, I’d spend 55 minutes thinking about the problem and 5 minutes thinking about solutions.” The importance of clearly defining the question prior to starting a review cannot be emphasized enough. The exact question asked will determine the inclusion/exclusion of the studies in review, affecting the ultimate findings and potential generalizability of results to clinical settings.

In EBM, we typically define a structured question in “PICO” format: population, intervention, comparator, and outcome. In addition, most reviews will define a minimum study type for inclusion. Whilst the gold standard is the randomized controlled trial (RCT), these are not always feasible or available, and it is possible (although perhaps less desirable) to perform a systematic review of nonrandomized studies. Once the question is defined, it can then be translated into a series of inclu-

sion/exclusion criteria and used to define the search strategy for relevant studies.

Step 2: Literature Search

The next step is to design a search strategy that will identify the relevant literature for the research question. Search strategies for systematic reviews tend to be broad, with the focus very much on sensitivity rather than specificity to avoid missing relevant studies. Combinations of subject headings/keywords and free text terms for the population, intervention, and possible outcomes of interest are used. Ideally, the use of language restrictions should be avoided if relevant articles can be translated.

Usually, searches are conducted across multiple databases to improve sensitivity. For surgical reviews, including RCTs, optimum databases usually include MEDLINE and Cochrane CENTRAL; whilst, Web of Science may replace CENTRAL for nonrandomized studies [10]. Reference lists of relevant articles and reviews can be searched for additional studies, and it is good practice to search clinical trial registers for unpublished and on-going studies.

Search results should be screened by 2 reviewers independently, to compare studies to the pre-defined inclusion/exclusion criteria. Initially, references are screened by title and abstract, and then those that are potentially relevant undergo a full text review. It is useful to record numbers of references and reasons for exclusion at each stage of the process for full transparency.

Step 3: Risk of Bias Assessment

Once the studies to be included in the review have been defined, the quality of the primary literature should be assessed. Studies with poor methodology and bias included in a review have a significant risk of introducing bias into the review findings. Where study quality is poor, meta-analysis is usually ill-advised, and at very least a sensitivity analysis of higher quality studies should be undertaken.

There are a wide variety of quality assessment tools in use, but the general principles of all are the same. The tools assess for risk of bias in domains such as subject selection, allocation, treatment administration, outcome measurement, attrition, and outcome reporting. Most commonly used for RCTs are the Cochrane risk of bias tool and the Jadad scale [11, 12], although it should be noted that the Jadad scale misses important concepts such as use of intent-to-treat analysis and assessment for adequate allocation concealment. Down and Black [13] is popular for nonrandomized studies.

Step 4: Data Extraction and Synthesis

The approach to this step depends very much on the amount and quality of the primary evidence identified. Where evidence is of poor quality or sparse, it is likely that a narrative description looking for trends and gaps in the data will be most appropriate. Where data are sufficient, meta-analysis can be attempted.

Data extraction should be performed independently by 2 researchers to reduce the risk of errors. For binary outcomes (e.g., anastomotic leak rates), the number of patients experiencing the outcome, and the denominator for each arm of each study are required. For continuous endpoints (e.g., serum creatinine level), a mean, standard deviation and number of patients for each arm are required.

Meta-analysis involves the combination of effect sizes for each included study to provide a weighted overall analysis. Commonly used statistical models are either “fixed effects” or “random effects”. A fixed effects model assumes 1 common effect of the intervention across all studies. A random effects model accepts that there is likely to be a randomly (usually normally) distributed variation in effect between different study populations, resulting in a more conservative estimate of effect variability. Whilst in practice reviewers will often use a fixed effects analysis in the absence of significant between-study heterogeneity, some authors have argued that the inherent variability in surgical practice and techniques means that a random effects analysis should always be used in reviews of surgical interventions [14].

If meta-analysis is employed, the reviewer should always assess for, and attempt to explain, any between-study heterogeneity [15]. This is usually assessed by means of the I^2 statistic or Cochran Q test and attempts to identify between-study heterogeneity that cannot be explained by chance alone. Where there is significant heterogeneity (typically a Cochran p value <0.1 , or $I^2 > 50\%$), the included studies should be reviewed for differences in the population, intervention, or outcome definition that may explain the different effect sizes seen. Visual inspection of the forest plot, subgroup analyses, and meta-regression can all be useful in exploring causes. Where no obvious cause is identified; meta-analyses should be interpreted with caution.

As described above, a meta-analysis is only as good as the included studies. It is well described that negative studies are less likely to be published due to lack of interest, commercial pressures and editorial biases, and this creates problems for the systematic reviewer [16–18]. Reviews of published studies are more likely to include positive findings and thus favour the intervention – so-called “publication bias.” This can also happen at an individual outcome level – restriction on word count for journal articles and lack of interest will mean that for a given study, statistically significant outcomes are more likely to be reported than non-significant ones. This again leads to bias in meta-analysis.

Attempts to reduce risk of publication bias can include searching for unpublished studies and contacting authors for missing outcomes. Where it is not possible to remove bias, the magnitude of the problem can at least be assessed by the use of techniques such as funnel plots, Egger’s test, and trim and fill [19]. However, this can be challenging with small numbers of studies.

Step 5: Interpretation

Once analyses are complete, the findings must be interpreted in light of the quality and quantity of the underlying evidence and both statistical and clinical significance of the findings. Meta-analysis of large numbers of studies can lead to findings that whilst statistically significant, are of limited clinical significance. Alternatively, a lack of evidence can mean that despite meta-analysis there is still uncertainty as to the value of an intervention.

It is important to interpret the potential impact of the evidence in a real-world setting. Clinical trials often have restrictive inclusion/exclusion criteria and are able to show benefits to interventions that may work well in a supervised trial setting, but logistical or cost constraints lead to failure in real-world practice. When reading a systematic review,

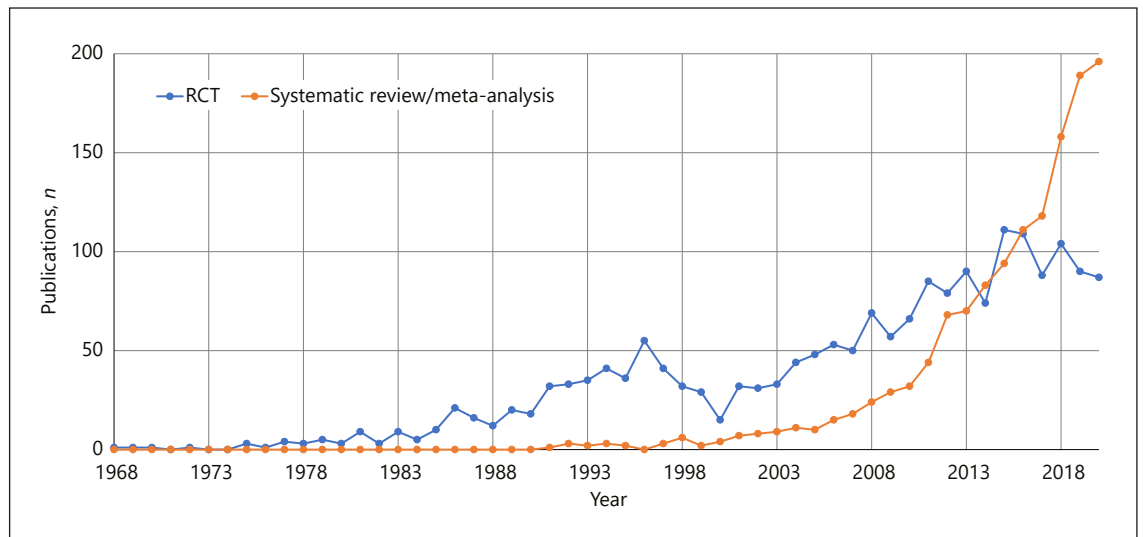


Fig. 1. Trends in publication of RCTs and systematic reviews in surgery 1968–2020. Data from PubMed (www.ncbi.nlm.nih.gov/pubmed), using the MeSH heading “Specialties, surgical.” RCT, randomized controlled trial.

the reader must ask whether the patients and interventions in the underlying studies are similar enough to their own practice that the results are likely to be replicated, and whether delivery of the intervention will be feasible in the setting of their own hospital and healthcare system.

Limitations of Systematic Reviews

There has been an explosion in systematic reviews of surgical topics over the past decade, to the point where multiple reviews of the same topic have been published with a great deal of redundancy and sometimes conflicting results [20, 21]. In fact, data from MEDLINE suggest that the number of new systematic reviews of surgical topics published each year now outnumbers the number of new RCTs (Fig. 1). The old adage “garbage in, garbage out” holds very true here: the paucity of RCTs means that many of these reviews resort to synthesis of small numbers of observational studies with a high risk of bias. This limits the validity and scope of the conclusions drawn.

Sometimes reviewers themselves may have vested interests. Pharmaceutical or medical device companies will sometimes commission reviews of topics that support their interests, and there may be pressure on the review authors to interpret the data in a favourable light. Bias can be introduced by introducing or selecting specific inclusion/exclusion criteria to limit studies included in the review, or by the selective outcome reporting that is prob-

lematic in many clinical trials. There is good evidence to suggest that industry-funded reviews are less transparent, less critical of limitations in the included clinical trials, and more likely to have favourable conclusions than independent reviews of the same products [22]. Some of these issues can be avoided by protocol registration and robust, transparent methodology.

Often, in a desire to undertake a meta-analysis of a meaningful number of studies, reviewers are tempted to include very heterogeneous study populations with differences in the populations and interventions that may impact the clinical effectiveness. This will lead to heterogeneity in the meta-analysis, and one should always be suspicious of a review with significant unexplained heterogeneity.

Challenges Specific to Systematic Reviews of Surgical Interventions

Many of the challenges in conducting systematic reviews of surgical interventions reflect the same issues in conducting the underlying surgical clinical trials. Because of the unique difficulties faced when conducting good quality trials in surgery, RCTs are often lacking and established historical practice is often unchallenged or based on observational studies. The quality of trials is slowly improving over time, but still has some way to go [3].

Heterogeneity in Surgery

There are many aspects of patient care that contribute to successful outcome following surgery. Differences in pre-operative assessment and management, surgical technique and expertise, anaesthetic, and post-operative care (including use of enhanced recovery programmes) can all impact the chances of successful outcome. In any 1 trial, surgical procedures may be performed by a number of different surgeons in different institutions, each with their own preferences and experience levels. This has been documented in trials of radical prostatectomy, with between-surgeon variations contributing significantly to heterogeneity in outcomes [23]. For this reason, results of clinical trials reporting to be of the same surgical intervention may be impacted by differences in surgical technique and concomitant care, which will result in heterogeneity in meta-analysis. The systematic reviewer must carefully examine all aspects of the patients' perioperative care and the exact nature of the surgical procedure when attempting to explain sources of heterogeneity.

Surgeon Preference and Learning Curves

Comparison of different surgical interventions, or comparison of surgical and medical interventions, can be made difficult by issues with surgeon or institutional availability or preference. Surgeons often have greater experience with a particular procedure or method, which introduces the problem of familiarity and learning curve [24, 25]. Asking a surgeon to conduct an unfamiliar new procedure will inevitably have worse results initially, with longer operative times and perhaps more bleeding and higher complication rates. As the surgeon gains experience, results will improve until a steady state is achieved. However, if a trial is conducted beyond the end of the learning curve, the surgeon may have lost equipoise and be unwilling to randomize to an alternative treatment. Conducting a trial during the learning curve may lead to unfavourable comparisons to established treatments, and inclusion of early studies of a new intervention in a systematic review may affect the size of effects seen in meta-analysis.

Patient Preference

Patients will also often have a preference for a type of treatment or procedure. Surgical treatments tend to have a more immediate short-term impact, trading greater perioperative risk with the promise of more immediate or permanent results or relief of symptoms than nonsurgical alternatives. In conditions such as acute appendicitis where surgical risk is relatively low, this may lead to a

preference for surgical intervention [26]. In conditions where the surgical option could have a greater impact on patient quality of life, such as risk of -ostomy following surgery for ulcerative colitis, patients may be willing to accept greater longer term risk from medical therapy [27]. Such preferences can make recruitment to surgical trials difficult, and lead to imbalance in patient characteristics within a trial due to crossover and noncompliance. Again, the reviewer must look for evidence of this when including studies in a systematic review.

Blinding

One of key components of any assessment of risk of bias is evidence of blinding of participants and investigators. Blinding reduces the risk of performance bias (treating participants differently depending on their allocated group) and measurement bias (differential outcome assessment, particularly of subjective or patient-reported outcomes). Blinding of surgical interventions can be challenging and often impossible, and so it is very important that the reviewer looks for evidence of bias. Use of objective endpoints and blinding of outcome assessment (e.g., histopathology and radiology) can help to reduce the risk. Analysis of RCTs in surgery has demonstrated that blinding could be used much more frequently, and lack of blinding leads to over-estimation of treatment effects [28]. Rather than using the term "double-blind" for surgical trials, it is perhaps more useful to consider whether each stakeholder (patient, surgeon, outcome assessor, and statistician) could have been blinded, and whether this blinding was undertaken.

Emergency Interventions

Many surgical procedures are undertaken in the emergency setting, which provides additional challenges for the ethics and recruitment to clinical trials. Patients in the emergency setting may not have the capacity to consent, and timescales to consider recruitment prior to intervention will be very limited. Recruitment to studies is often pragmatic, based upon early assessment prior to formal diagnosis, which may lead to more crossovers and later exclusions. A good example of this is the IMPROVE study of endovascular repair of ruptured aortic aneurysm [29]. Patients were recruited following clinical assessment by a senior emergency-room doctor, prior to imaging confirmation of a diagnosis of ruptured aneurysm. This inevitably led to some exclusions due to misdiagnosis, and further crossovers due to patient/surgeon preference, availability of endovascular repair, and suitability for an endovascular approach. Whilst this is a perfectly accept-

able and pragmatic way to facilitate recruitment in the emergency setting, the results need to be scrutinized carefully to look for introduction of bias in the treated groups that may impact the outcome.

Industry Involvement

Developing an evaluation of new surgical techniques and equipment can be expensive, and as such often requires the involvement of industry to support funding. As with other fields of medicine, there is evidence that industry involvement in surgical RCTs can bias outcomes. Probst and colleagues evaluated 165 RCTs in general and abdominal surgery, of which 115 were industry sponsored [30]. Industry-sponsored studies had equivalent methodological quality to nonsponsored studies but were significantly more likely to report a positive outcome (odds ratio 5.79). It is therefore important that source of funding should be evaluated in systematic review, and subgroup analysis by funding source considered when evaluating surgical interventions.

Conclusions

After many years of lagging behind other areas of medicine, the number and quality of clinical trials in surgical specialities are increasing. This has led to an explosion in the number of systematic reviews of surgical topics. Whilst these reviews can be useful in creating evidence-based practice guidelines, it is important that they are conducted as rigorous scientific studies using robust

methodology and paying particular attention to the scope and quality of the underlying evidence. Surgical clinical trials pose unique challenges that must be considered in the review process. Often a review will identify more questions than answers, and this in itself is valuable in identifying evidence gaps and guiding the direction and design of future primary research.

Well-conducted systematic reviews represent the highest level of evidence upon which we can guide clinical practice. The improvements in the quality of RCTs and systematic reviews in surgery are welcomed, and hopefully further improvements in the future will allow us to ensure that delivery of surgical interventions will share an equal quality evidence base with other specialities.

Conflict of Interest Statement

Simon R. Knight has previously received consultancy fees for clinical trial design from OrganOx Ltd. There are no conflicts of interest to declare in relation to the content of this review.

Funding Sources

No external funding was received for the conduct of this review.

Author Contributions

This invited review was researched and written entirely by Simon R. Knight.

References

- 1 Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996 Jan 13;312(7023):71–2.
- 2 OCEBM Levels of Evidence Working Group. *The Oxford levels of evidence 2 [Internet]*. 2011 [cited 2021 Aug 17]. Available from: <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebml-levels-of-evidence>.
- 3 Yu J, Li X, Li Y, Sun X. Quality of reporting in surgical randomized clinical trials. *Br J Surg*. 2017 Feb;104(3):296–303.
- 4 Yu J, Chen W, Wu P, Li Y. Quality of reporting of systematic reviews and meta-analyses of surgical randomized clinical trials. *BJS Open*. 2020 Jun;4(3):535–42.
- 5 Meshikhes AW. Evidence-based surgery: the obstacles and solutions. *Int J Surg*. 2015 Jun 1; 18:159–62.
- 6 Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021 Mar 29;372:n71.
- 7 Hughes ES. Asepsis in large-bowel surgery. *Ann R Coll Surg Engl*. 1972 Dec;51(6):347–56.
- 8 Güenaga KF, Matos D, Castro AA, Atallah AN, Wille-Jørgensen P. Mechanical bowel preparation for elective colorectal surgery. *Cochrane Database Syst Rev*. 2011 Sep 7(9): CD001544.
- 9 Gustafsson UO, Scott MJ, Hubner M, Nygren J, Demartines N, Francis N, et al. Guidelines for perioperative care in elective colorectal surgery: Enhanced Recovery After Surgery (ERAS®) society recommendations: 2018. *World J Surg*. 2019 [cited 2021 Aug 17];43(3): 659–95. Available from: https://click.endnote.com/viewer?doi=10.1007%2Fs00268-018-4844-y& token=WzUwMzA3NSwiMTAuMTAwNy9zMDAyNjgtMDE4LTQ4NDQt eSjd.00DfbPgWCLRvZT5dEaBDv_56Y_0.
- 10 Goossen K, Tenckhoff S, Probst P, Grummich K, Mihaljevic AL, Büchler MW, et al. Optimal literature search for systematic reviews in surgery. *Langenbecks Arch Surg*. 2018 Feb; 403(1):119–29.
- 11 Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996 Feb;17(1):1–12.
- 12 Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011 Oct 18;343: d5928.

- 13 Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998 Jun;52(6):377–84.
- 14 Kalkum E, Klotz R, Seide S, Hüttner FJ, Kowalewski K-F, Nickel F, et al. Systematic reviews in surgery: recommendations from the Study Center of the German Society of Surgery. *Langenbecks Arch Surg*. 2021 Jun 15 [cited 2021 Aug 17]. Epub ahead of print.
- 15 Hatala R, Keitz S, Weyer P, Guyatt G. Tips for learners of evidence-based medicine: 4. Assessing heterogeneity of primary studies in systematic reviews and whether to combine their results. *CMAJ*. 2005 Mar 1;172(5):661–5.
- 16 Dimick JB, Diener-West M, Lipsett PA. Negative results of randomized clinical trials published in the surgical literature: equivalency or error? *Arch Surg*. 2001 Jul 1;136(7):796–800.
- 17 Duyx B, Urlings MJE, Swaen GMH, Bouter LM, Zeegers MP. Scientific citations favor positive results: a systematic review and meta-analysis. *J Clin Epidemiol*. 2017 Aug;88:92–101.
- 18 Fanelli D. Negative results are disappearing from most disciplines and countries. *Scientometrics*. 2011 Sep 11;90(3):891–904.
- 19 Sterne JA, Egger M, Smith GD. Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. *BMJ*. 2001 Jul 14;323(7304):101–5.
- 20 Mascarenhas R, Chalmers PN, Sayegh ET, Bhandari M, Verma NN, Cole BJ, et al. Is double-row rotator cuff repair clinically superior to single-row rotator cuff repair: a systematic review of overlapping meta-analyses. *Arthroscopy*. 2014 Sep;30(9):1156–65.
- 21 Katsura M, Kuriyama A, Tada M, Yamamoto K, Furukawa TA. Redundant systematic reviews on the same topic in surgery: a study protocol for a meta-epidemiological investigation. *BMJ Open*. 2017 Aug;7(8):e017411.
- 22 Jørgensen AW, Hilden J, Gøtzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. *BMJ*. 2006 Oct 12;333(7572):782.
- 23 Nyberg M, Sjöberg DD, Carlsson SV, Wilderäng U, Carlsson S, Stranne J, et al. Surgeon heterogeneity significantly affects functional and oncological outcomes after radical prostatectomy in the Swedish LAPPRO trial. *BJU Int*. 2021;127(3):361–8.
- 24 Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning curve effects in surgical trials. *Clin Trials*. 2004;1(5):421–7.
- 25 Papachristofi O, Jenkins D, Sharples LD. Assessment of learning curves in complex surgical interventions: a consecutive case-series study. *Trials*. 2016 Jun 1;17(1):266.
- 26 Hanson AL, Crosby RD, Basson MD. Patient preferences for surgery or antibiotics for the treatment of acute appendicitis. *JAMA Surg*. 2018 May 1;153(5):471–8.
- 27 Bewtra M, Kilambi V, Fairchild AO, Siegel CA, Lewis JD, Johnson FR. Patient preferences for surgical versus medical therapy for ulcerative colitis. *Inflamm Bowel Dis*. 2014 Jan; 20(1):103–14.
- 28 Probst P, Zschke S, Heger P, Harnoss JC, Hüttner FJ, Mihaljevic AL, et al. Evidence-based recommendations for blinding in surgical trials. *Langenbecks Arch Surg*. 2019 May 1;404(3):273–84.
- 29 IMPROVE Trial Investigators; Powell JT, Sweeting MJ, Thompson MM, Ashleigh R, Bell R, Gomes M, et al. Endovascular or open repair strategy for ruptured abdominal aortic aneurysm: 30 day outcomes from IMPROVE randomised trial. *BMJ*. 2014 Jan 13;348:f7661.
- 30 Probst P, Knebel P, Grummich K, Tenckhoff S, Ulrich A, Büchler MW, et al. Industry bias in randomized controlled trials in general and abdominal surgery: an empirical study. *Ann Surg*. 2016 Jul;264(1):87–92.