



Systematic reviews in surgery—recommendations from the Study Center of the German Society of Surgery

Eva Kalkum¹ · Rosa Klotz^{1,2} · Svenja Seide³ · Felix J. Hüttner⁴ · Karl-Friedrich Kowalewski⁵ · Felix Nickel² · Elias Khajeh² · Phillip Knebel^{1,2} · Markus K. Diener¹ · Pascal Probst^{1,2} 

Received: 7 May 2021 / Accepted: 16 May 2021 / Published online: 15 June 2021
© The Author(s) 2021

Abstract

Background Systematic reviews are an important tool of evidence-based surgery. Surgical systematic reviews and trials, however, require a special methodological approach.

Purpose This article provides recommendations for conducting state-of-the-art systematic reviews in surgery with or without meta-analysis.

Conclusions For systematic reviews in surgery, MEDLINE (via PubMed), Web of Science, and Cochrane Central Register of Controlled Trials (CENTRAL) should be searched. Critical appraisal is at the core of every surgical systematic review, with information on blinding, industry involvement, surgical experience, and standardisation of surgical technique holding special importance. Due to clinical heterogeneity among surgical trials, the random-effects model should be used as a default. In the experience of the Study Center of the German Society of Surgery, adherence to these recommendations yields high-quality surgical systematic reviews.

Keywords Synoptic evidence · Systematic review · Surgery · Meta-analysis · Evidence-based medicine

Background

Systematic reviews (SRs) are of high importance for decision-makers in the healthcare system and crucial to the

development of clinical guidelines. SRs connect the results of single studies on the same topic, thereby providing clinicians with the best foundation for evidence-based treatment of their patients. Additionally, SRs can identify research gaps

✉ Pascal Probst
pascal.probst@med.uni-heidelberg.de;
pascal.probst@swissonline.ch

Eva Kalkum
eva.kalkum@med.uni-heidelberg.de

Rosa Klotz
rosa.klotz@med.uni-heidelberg.de

Svenja Seide
seide@imbi.uni-heidelberg.de

Felix J. Hüttner
felix.huettner@uniklinik-ulm.de

Karl-Friedrich Kowalewski
karl-friedrich.kowalewski@umm.de

Felix Nickel
felix.nickel@med.uni-heidelberg.de

Elias Khajeh
elias.khajeh@med.uni-heidelberg.de

Phillip Knebel
phillip.knebel@med.uni-heidelberg.de

Markus K. Diener
markus.diener@med.uni-heidelberg.de

¹ Study Center of the German Society of Surgery (SDGC), University of Heidelberg, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany

² Department of General, Visceral and Transplantation Surgery, University of Heidelberg, Im Neuenheimer Feld 420, 69120 Heidelberg, Germany

³ Institute of Medical Biometry and Informatics, University of Heidelberg, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany

⁴ Department of General and Visceral Surgery, University Hospital Ulm, Albert-Einstein-Allee 23, 89081 Ulm, Germany

⁵ Department of Urology and Urological Surgery, University Medical Center Mannheim, University of Heidelberg, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Germany

and provide recommendations for future clinical trials in terms of effect estimates and meaningful endpoints.

Generally speaking, a systematic review has five steps: *formulating* the research question; *identifying*, *selecting*, and *assessing* the relevant literature; and *synthesis*, i.e. interpretation of quantitative results (such as by meta-analysis) in light of the quality of the studies included [1]. SRs are considered original research in most journals, especially when a meta-analysis is performed [2, 3].

Only a limited number of surgical interventions are based on randomised controlled trials (RCTs) [4, 5] representing the highest level of evidence [6]. Therefore, surgical SRs are especially at risk of the falling prey to the classic ‘garbage in, garbage out’ problem [7]. To avoid this designation, SRs in the field of surgery in particular must address the quality of all included trials, paying special attention to specifics about surgical trial methodology. Otherwise, poor-quality SRs run the risk of encouraging poor treatment decisions and incurring unnecessary costs within the healthcare system [8].

In 2005, the Study Center of the German Society of Surgery (www.sdgc.de) founded a systematic review working group. This working group is committed to disseminating the know-how required to plan, conduct, and publish SRs among German surgeons and to aid them throughout this process. Since 2005, the group published more than 70 SRs and specific literature was created on the methodology of surgical SRs [9–12]. In this article,

recommendations for conducting a state-of-the-art surgical SR with or without meta-analysis are provided.

Recommendations

Recommendations are given for each step of an SR. General recommendations are followed by specific recommendations important to surgical reviews.

Formulating the research question

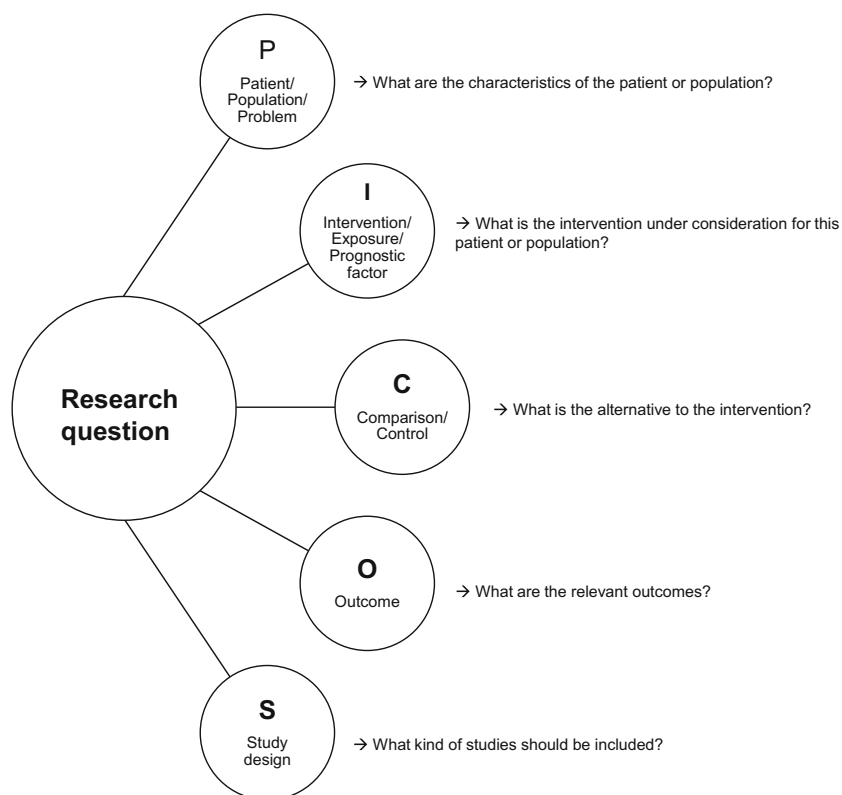
At the beginning of an SR, it is essential to clearly state the research question. The aim of an SR should be to answer an unanswered and important clinical question without an existing SR or for which new primary evidence has become available. Therefore, before beginning an SR, registers and literature databases should be screened for existing SRs on the subject.

To define a well-focused and answerable research question, PICOS criteria [13, 14] should be used. PICOS stands for Patient, Intervention, Comparison, Outcome, and Study design (Fig. 1). In surgical reviews, the following specific characteristics should be clarified:

Patient:

- Will only adult patients be included?

Fig. 1 PICOS criteria



- Will only patients with a malignant disease be considered?
- Will patients with previous surgery be excluded?
- Will patients with neoadjuvant therapy be excluded?
- etc.

Intervention:

- Due to the complexity of surgical interventions, it is important to define the exact procedure or group of relevant procedures to guarantee comparability and to ensure that if interventions are deemed effective, they can actually be reproduced and implemented in clinical practice.
- In so doing, the recommendations of Blencowe et al. [15] should be followed. Their paper describes a framework for deconstructing surgical interventions into their constituent components and provides steps to clarify details of the intervention under evaluation (e.g. concerning the intervention ‘robotic partial pancreatoduodenectomy’; this list could include positioning of the robot, incision and access, dissection, reconstruction, anastomoses, and closure).

Comparison: adequate controls in surgical reviews can be any of the following, depending on the research question:

- One or more state-of-the-art surgical intervention(s) (‘gold standard’, e.g. intervention: laparoscopic pancreatoduodenectomy versus comparison: open pancreatoduodenectomy [16])
- Conservative treatment without surgical intervention such as the administration of a drug (e.g. intervention: metabolic surgery versus comparison: medical treatment in patients with type 2 diabetes [17])
- Surgical placebo [18, 19] (e.g. intervention: arthroscopic debridement of the knee versus comparison: sham surgery, i.e. skin incision only in patients with osteoarthritis [20])

Outcome: a main outcome should be defined and clear definitions should be applied to it. Common outcomes in surgical reviews include the following:

- Intraoperative outcomes, such as operative time or blood loss
- Perioperative morbidity and mortality and specific post-operative complications
- Patient-reported outcome measures, such as functional outcomes and quality of life
- Long-term oncological or functional outcomes
- Economic evidence, such as cost-effectiveness

Study design: depending on the available literature, it has to be decided what study type(s) should be included:

- RCTs
- Non-randomised prospective (comparative) studies
- Cohort-type studies
- Case series

After formulating the research question, a research protocol should be developed. It is recommended to register the SR in a public register, e.g. PROSPERO [21, 22], and/or to publish the protocol in a peer-reviewed journal. Registration improves the quality of performance and the transparency in an SR, prevents duplicate work, and reduces the risk of selective reporting by providing an a priori analysis plan. The PRISMA-P checklist should be followed when writing an SR protocol [23].

Identifying potentially relevant literature

The aim of a systematic literature search is to identify all relevant studies regarding the research question without including any irrelevant studies at the same time. However, this is impossible: a too-narrow or highly specific search strategy will miss relevant studies, whereas a too-broad search strategy would require screening thousands of irrelevant studies.

It is recommended to start with a wide initial scoping search and to adjust the search accordingly. Depending on the number and content of studies retrieved, the literature search should then be modified. It is important to include all free text synonyms of a term, e.g. ‘Whipple’, ‘pancreatoduodenectomy’, ‘pancreaticoduodenectomy’, ‘resection of the pancreatic head’, and ‘pancreatic head surgery’. Additionally, using medical subject headings (MeSH) improves the quality of the literature search. If there are still too many hits, adjuvant filters, such as a time window (e.g. time of the first available robotic surgery) or study type (e.g. RCT filter), can be considered. After the initial search, terms can be revised according to whether or not relevant, previously known literature can be found using these search terms.

An SR should always search more than one database. The following databases are recommended for surgical SRs: MEDLINE (via PubMed), Web of Science, and Cochrane Central Register of Controlled Trials (CENTRAL). For surgical topics, MEDLINE has the highest recall (92.6%) and precision for non-randomised studies (NRS, 5.2%), whereas CENTRAL is more sensitive (88.4%) and has the highest precision (8.3%) for RCTs. The combination of MEDLINE and CENTRAL has a 98.6% recall for RCTs. For NRS, the highest recall (99.5%) is retrieved by the combination of MEDLINE and Web of Science [9]. EMBASE does not contribute substantially to reviews on a surgical intervention. However, for research questions involving a drug intervention (e.g. medical vs. surgical intervention), the inclusion of EMBASE should be considered [9]. An additional hand search is also recommended; reference lists of relevant articles found by the literature search should be screened for further

relevant articles not found by the literature search. It is also advisable to seek professional assistance by a librarian if the research team has limited experience in conducting a search.

Selecting the relevant literature

The most relevant inclusion and exclusion criteria for study selection have already been outlined by PICOS. However, these a priori stated criteria could sometimes be altered depending on the number of eligible studies. Besides the clinically based inclusion and exclusion criteria, the authors must determine which study designs should be included. Although the quantity and quality of surgical RCTs have increased in some specialities [24], there is still a general lack of high-quality trials in many areas of surgery. Therefore, the aim of an SR should be to find the best available evidence. If four or more RCTs are available, then NRS can be omitted.

Limiting oneself to articles in English should be avoided and endeavours should be made to translate articles in other languages. If non-English articles are excluded, the exact number of those articles should be provided.

Overall, the final eligibility criteria should be stated clearly, with the publication of results. While one reviewer can conduct the initial screening of titles and abstracts, it is recommended that both initial screening and full-text screening (according to eligibility criteria) be completed by two researchers [13]. Any disagreement during the screening process should be resolved by consensus, or by consultation with a third reviewer. The selection process should be documented with a PRISMA flowchart (http://www.prisma-statement.org/documents/PRISMA_2020_flow_diagram_new_SRs_v2.docx) [22]. In addition, during full-text screening, the reason for exclusion should be given.

Assessing the quality of primary studies

Studies meeting all inclusion criteria enter the next stage of data extraction. This step should be performed using a standardised form (electronic or paper-based), which was piloted in initial trials and revised accordingly. Besides the extraction of the relevant endpoints, some descriptive details should be presented in a tabular form in the publication, e.g. author, title, year and geographical origin of publication, number of participants, and specific information on the surgery performed. Additionally, baseline information such as tumour stage or performance status could be included when some specialised centres perform surgery on more complex cases, thus potentially impacting the outcome. It is advisable to involve a statistician at this stage, in order for data to be extracted in a manner which allows for the calculations planned.

Outcomes should be predefined before screening the eligible literature and should refer to PICOS. A main outcome, such as the primary endpoint in a trial, should be clearly defined. For easier comparison among outcomes, clearly defined

endpoints should be used whenever possible. For example, postoperative complications should be evaluated according to the Clavien-Dindo classification [25] and endpoints like postoperative pancreatic fistula should be extracted according to ISGPS definitions [26]. The use of differing definitions among the studies included should be mentioned, and their possible impact on quantitative analysis should be discussed (and could be assessed) in a sensitivity analysis.

Assessing a study's methodological quality is a key aspect of every well-performed SR. Only in light of methodological quality can the quantitative merit of a study be interpreted. This aspect is even more important for surgical SRs, due to the lack of standardisation in surgical trials compared to pharmacological trials. Critical appraisal of the studies with a validated tool is therefore also recommended. Risk of bias should be reported in a paragraph dedicated to it in the results section. Different risk of bias assessment tools exist for different study types (see Table 1 for recommendations).

Some other tools which have been critically dealt with, especially for reporting bias, exist for different types of studies [32].

The critical appraisal of surgical trials has many specifics that should be addressed, including but not limited to blinding, industry bias, experience, and standardisation of intervention.

Generally, in RCTs, blinding is favourable to reduce detection and performance bias. In surgical trials, blinding is not easy to apply [33, 34] and not every lack of blinding will lead to performance or detection bias. The usual term 'double-blind trial' is not reasonably transferable to surgical trials and should be avoided. Therefore, it is recommended to report specifically which (if any) study contributor (whether patient, surgeon, data collector, outcome assessor, or data analyst) was blinded.

Specifically, for surgical trials, it has been shown that industry funding leads to more positive results than independent funding [11]. Therefore, in surgical trials, which investigate an intervention with an inherent industrial interest, the source of funding should be extracted and subgroup analysis of industry-funded and non-industry-funded trials should be performed [35].

Furthermore, experience of the operating surgeon(s) and possible learning curves should be addressed, since these can specifically influence comparisons between established and new surgical interventions [36, 37]. Validated surgical

Table 1 Risk of bias assessment tools

Tool	Study type
RoB 2.0 [27]	RCTs
ROBINS-I [28]	NRS (clinical controlled trial, case-control)
MINORS [29]	NRS
Downs and Black [30]	NRS (case series)
QUIPS [31]	Prognostic studies

quality control measures should be implemented since significant bias can otherwise result [16, 38].

Small-study effects and publication bias

Small-study effects describe the phenomenon that smaller studies sometimes show different treatment effects than larger ones [39]. The most well-known (albeit not the only) reason for this phenomenon is the presence of publication bias in the data. This bias occurs if the chance of a smaller study being published increases when it shows a stronger treatment effect, which in turn biases the results of the meta-analysis and the SR [40]. Small-study effects can be graphically illustrated by a funnel plot [41], where estimated treatment effects are shown against a measure of their precision. In the absence of small-study effects, the funnel plot shows a symmetric scattering of the treatment effects around their average in the form of a triangle, with more variation in smaller (imprecise) studies than in larger (precise) ones. There are several well-known statistical tests for measuring small-study effects and asymmetry in funnel plots, e.g. the non-parametric Begg and Mazumdar test [42] or the parametric Egger regression test [43]. However, the power of statistical tests is known to be low, and interpretability of funnel plots is often limited, due to the low number of studies contributing to the analysis and should therefore only be used when more than ten studies are available [44].

Finally, surgical interventions are complex and comprise multiple components that might be accompanied by concomitant interventions such as anaesthesia and perioperative management [45]. Also, different levels of experience among operating surgeons as well as case volumes among the hospitals where a trial was performed might influence outcomes. For example, an SR addressing how details about surgical interventions are reported found a clear standardisation of the intervention in less than 30% of the included RCTs [46]. Also, measurements of adherence to the intervention might be missing. Consequently, the level of standardisation of the intervention in the included trials needs to be reported and in case of missing standardisation in some trials, a subgroup analysis including only trials containing clear reporting of the performed intervention might be necessary.

Contrary to screening, data extraction and risk of bias assessment should be performed by at least two researchers [13] and any differences among them should be resolved by consensus, or by consultation with a third reviewer.

Furthermore, whether or not any measures of surgical experience were gathered in the primary trials, as well as its impact on outcome, should also be assessed in a sensitivity analysis if such information is available.

The results of the critical appraisal should be clearly described in the final report of any SR, and the impact of the quality of evidence on interpretation of the results should be discussed. For this reason, simply providing a summary score

of quality assessment on the study level is strongly discouraged.

Synthesis of quantitative and qualitative results

The last step of an SR is to synthesise the extracted data, i.e. quantitative analysis (meta-analysis), and merge it with qualitative analysis (critical appraisal/risk of bias).

A quantitative analysis should always be evaluated critically, and determining whether or not a meta-analysis makes sense should be based on the extracted data. A meta-analysis is justified if sufficiently homogeneous studies can be assimilated for statistical analysis. Different effect measures are used to summarise treatment effects depending on the scale of the outcome. Commonly used effect measures are the risk difference (RD), the risk ratio (RR) and the odds ratio (OR) for binary outcomes, the mean difference (MD) or standardised mean difference (SMD) for continuous outcomes, and hazard ratios (HR) for time-to-event outcomes [47]. When choosing adequate effect measures for the included endpoints, statistical aspects, as well as convention or interpretability, need to be considered [48].

A meta-analysis summarises the results of individual studies as a combined effect estimate. As studies usually differ in the number of patients included and therefore vary in their precision, the naïve approach of simply averaging the effects across studies is not recommended. Instead, weighted effect estimators that use the precision of the identified studies are commonly used to include the individual studies according to their precision and more weight will be assigned to large, precise studies than to small, imprecise ones [49]. Analytical techniques can be broadly classified into two categories: the fixed-effect (now often called common-effect) model and the random-effects model. The fixed-effect model assumes that all studies would yield the same result if they were infinitely large. Statistically, this means that random error is assumed to be attributable solely to the differences that occur in patients within a study and not due to any variations among the trials [50]. As a rule, this assumption is unrealistic, as small variations in study design and surgical technique almost always occur. In the random-effects model, such between-trial heterogeneity is considered, which then increases the imprecision with which the combined effect is estimated in a random-effects meta-analysis, i.e. leading to wider confidence intervals. Therefore, the use of the random-effects model is generally recommended in surgical SRs with meta-analysis irrespective of the presence of statistical heterogeneity as explained below [51]. Different estimation approaches are available. Well-established approaches include (upon others) the method of moment estimator of DerSimonian-Laird and the restricted maximum likelihood estimator [52–54]. A typical graphical

illustration of a meta-analysis is the forest plot. In a forest plot, the results of individual studies (estimated effects and confidence intervals) are displayed along with the results of the meta-analysis, the estimated between-trial heterogeneity, and the weight that is assigned to each study. It is recommended to calculate and display statistical heterogeneity by reporting between-trial variance as τ^2 and the I^2 as relative measure incorporating between- and within-trial variances [55, 56]. If this is unavoidable, e.g. due to a small number of included studies, the forest plot should be stratified or undergo sub-group analysis.

A variety of statistical approaches exist and determining which method is ‘best’ in a specific setting is often a matter of some debate. Therefore, the performance of sensitivity analyses to test the robustness of results is strongly advised. Commonly performed sensitivity analyses include a change of estimation method, an analysis with high-quality studies only, or an analysis including only studies with recent surgical methodology. Randomised and non-randomised studies should always be analysed separately.

Data synthesis should answer the research question by merging the qualitative and quantitative analyses.

Quantitative statements about outcomes, e.g. ‘operation A is superior to operation B’, should be accompanied by the certainty of evidence. For this step, the GRADE approach is recommended and certainty of evidence should be rated as very low, low, moderate, or high [57]. Apart from risk of bias, GRADE also includes other clinical and statistical characteristics which might influence the certainty of evidence. A ‘summary of findings’ table (<https://gradepro.org/>) showing quantitative results alongside the certainty of evidence for each outcome is therefore recommended. Moreover, since abstracts are read more frequently than full publications, the synthesis including certainty of evidence should be part of the abstract’s conclusion. Finally, PRISMA guidelines (<http://prisma-statement.org/>) should be followed when preparing and reporting SR results [22].

Summary

A thoroughly conducted SR of high-quality trials achieves the highest level of evidence. Ultimately, an SR of this calibre provides more comprehensive evidence for clinical decision-

Box 1 Summary of recommendations by the Study Center of the German Surgical Society for the conduct of a systematic review

Search the correct database! A search strategy for systematic reviews in surgery combining MEDLINE (via PubMed), Web of Science, and CENTRAL yields the highest level of precision and recall. EMBASE has no value for surgical interventions.

Appraise the study quality! A validated risk of bias assessment tool should be used. Additionally, some specifics of surgical trials should be noted:

- **Blinding** is also feasible in surgical trials and should be assessed carefully and reported in a surgical systematic review for every study contributor: patient, surgeon, data collector, outcome assessor, and data analyst.
- **Industry bias** is common in surgical trials if the investigated device poses an economic interest. Therefore, the source of funding should always be reported and considered as a potential source of bias.
- **Standardisation** is frequently unsatisfactory in surgical trials as surgical interventions are complex. Therefore, missing clear descriptions of the performed intervention, concomitant treatment, and hospital volume should be reported in systematic reviews and considered for the final conclusion. Further, the role of **experience/learning curves** in surgical performance and validated quality indicators should be addressed. This is especially important if new interventions are compared to established interventions since there is usually less experience with the newer intervention, leading to results favouring the established intervention.

Use a random-effects model for pooling! Clinical heterogeneity is always present between surgical trials—a sufficient reason to prefer the random-effects model to the fixed-effects model irrespective of statistical heterogeneity. Moreover, randomised and non-randomised trials should not be pooled.

Quality over quantity! Effects of low-quality trials are typically greater than those of high-quality trials. One of the reasons is bias. Therefore, quantitative results (meta-analysis) can only be interpreted in light of the quality of the trials included. Therefore, the core of a systematic surgical review is a critical appraisal of the literature.

making than a single study alone. An SR follows a structured process and requires specific methodology where surgical interventions are under investigation (Box 1).

Acknowledgements We thank Elizabeth Corrao-Billeter for language editing.

Authors' contributions All authors made substantial contributions to study conception and design and either drafted or revised the work. All authors gave their final approval to the publication of this manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Ethics approval and consent to participate For this article, no patients or animals were involved and an approval of an institutional review board was not necessary.

Conflict of interest The authors declare no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Khan KS, Kunz R, Kleijnen J, Antes G (2003) Five steps to conducting a systematic review. *J R Soc Med* 96(3):118–121. <https://doi.org/10.1258/jrsm.96.3.118>
- Krnic Martinic M, Meerpohl JJ, von Elm E, Herrle F, Marusic A, Puljak L (2019) Attitudes of editors of core clinical journals about whether systematic reviews are original research: a mixed-methods study. *BMJ Open* 9(8):e029704. <https://doi.org/10.1136/bmjopen-2019-029704>
- Meerpohl JJ, Herrle F, Reinders S, Antes G, von Elm E (2012) Scientific value of systematic reviews: survey of editors of core clinical journals. *PLoS One* 7(5):e35732. <https://doi.org/10.1371/journal.pone.0035732>
- Howes N, Chagla L, Thorpe M, McCulloch P (1997) Surgical practice is evidence based. *Br J Surg* 84(9):1220–1223
- Wente MN, Seiler CM, Uhl W, Buchler MW (2003) Perspectives of evidence-based surgery. *Dig Surg* 20(4):263–269. <https://doi.org/10.1159/000071183>
- Oxford Levels of Evidence Working Group (2011) The Oxford 2011 Levels of Evidence. <http://www.cebm.net/index.aspx?o=5653> [accessed 02/02/2021].
- Guller U (2008) Caveats in the interpretation of the surgical literature. *Br J Surg* 95(5):541–546. <https://doi.org/10.1002/bjs.6156>
- Shearman AD, Shearman CP (2012) How to practise evidence-based surgery. *Surgery (Oxford)* 30(9):481–485
- Goossen K, Tenckhoff S, Probst P, Grummich K, Mihaljevic AL, Buchler MW et al (2018) Optimal literature search for systematic reviews in surgery. *Langenbeck's Arch Surg* 403(1):119–129. <https://doi.org/10.1007/s00423-017-1646-x>
- Probst P, Huttner FJ, Klaiber U, Diener MK, Buchler MW, Knebel P (2015) Thirty years of disclosure of conflict of interest in surgery journals. *Surgery* 157(4):627–633. <https://doi.org/10.1016/j.surg.2014.11.012>
- Probst P, Knebel P, Grummich K, Tenckhoff S, Ulrich A, Buchler MW et al (2016) Industry bias in randomized controlled trials in general and abdominal surgery: an empirical study. *Ann Surg* 264(1):87–92. <https://doi.org/10.1097/SLA.0000000000001372>
- Probst P, Zschke S, Heger P, Harnoss JC, Huttner FJ, Mihaljevic AL et al (2019) Evidence-based recommendations for blinding in surgical trials. *Langenbeck's Arch Surg* 404(3):273–284. <https://doi.org/10.1007/s00423-019-01761-6>
- Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors) (2021) *Cochrane handbook for systematic reviews of interventions version 6.2* (updated February 2021). Available from www.training.cochrane.org/handbook.
- Richardson WS, Wilson MC, Nishikawa J, Hayward RS (1995) The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 123(3):A12–A13
- Blencowe NS, Mills N, Cook JA, Donovan JL, Rogers CA, Whiting P, Blazeby JM (2016) Standardizing and monitoring the delivery of surgical interventions in randomized clinical trials. *Br J Surg* 103(10):1377–1384. <https://doi.org/10.1002/bjs.10254>
- Nickel F, Haney CM, Kowalewski KF, Probst P, Limen EF, Kalkum E, Diener MK, Strobel O, Müller-Stich BP, Hackert T (2020) Laparoscopic versus open pancreaticoduodenectomy: a systematic review and meta-analysis of randomized controlled trials. *Ann Surg* 271(1):54–66. <https://doi.org/10.1097/SLA.0000000000003309>
- Billeter AT, Scheurle KM, Probst P, Eichel S, Nickel F, Kopf S, Fischer L, Diener MK, Nawroth PP, Müller-Stich BP (2018) Meta-analysis of metabolic surgery versus medical treatment for microvascular complications in patients with type 2 diabetes mellitus. *Br J Surg* 105(3):168–181. <https://doi.org/10.1002/bjs.10724>
- Probst P, Grummich K, Harnoss JC, Huttner FJ, Jensen K, Braun S et al (2016) Placebo-controlled trials in surgery: a systematic review and meta-analysis. *Medicine (Baltimore)* 95(17):e3516. <https://doi.org/10.1097/MD.0000000000003516>
- Beard DJ, Campbell MK, Blazeby JM, Carr AJ, Weijer C, Cuthbertson BH, Buchbinder R, Pinkney T, Bishop FL, Pugh J, Cousins S, Harris IA, Lohmander LS, Blencowe N, Gillies K, Probst P, Brennan C, Cook A, Farrar-Hockley D, Savulescu J, Huxtable R, Rangan A, Tracey I, Brocklehurst P, Ferreira ML, Nicholl J, Reeves BC, Hamdy F, Rowley SCS, Cook JA (2020) Considerations and methods for placebo controls in surgical trials (ASPIRE guidelines). *Lancet* 395(10226):828–838. [https://doi.org/10.1016/S0140-6736\(19\)33137-X](https://doi.org/10.1016/S0140-6736(19)33137-X)
- Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, Hollingsworth JC, Ashton CM, Wray NP (2002) A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 347(2):81–88. <https://doi.org/10.1056/NEJMoa013259>
- International Prospective Register of Systematic Reviews (PROSPERO) <https://www.crd.york.ac.uk/prospero/>.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA,

- Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:n71. <https://doi.org/10.1136/bmj.n71>
23. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M et al (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 4:1. <https://doi.org/10.1186/2046-4053-4-1>
 24. Huttner FJ, Capdeville L, Pianka F, Ulrich A, Hackert T, Buchler MW et al (2019) Systematic review of the quantity and quality of randomized clinical trials in pancreatic surgery. *Br J Surg* 106(1):23–31. <https://doi.org/10.1002/bjs.11030>
 25. Dindo D, Demartines N, Clavien PA (2004) Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 240(2):205–213. <https://doi.org/10.1097/01.sla.0000133083.54934.ae>
 26. Bassi C, Marchegiani G, Dervenis C, Sarr M, Abu Hilal M, Adham M, Allen P, Andersson R, Asbun HJ, Besselink MG, Conlon K, del Chiaro M, Falconi M, Fernandez-Cruz L, Fernandez-del Castillo C, Fingerhut A, Friess H, Gouma DJ, Hackert T, Izbicki J, Lillemoe KD, Neoptolemos JP, Olah A, Schulick R, Shrikhande SV, Takada T, Takaori K, Traverso W, Vollmer CR, Wolfgang CL, Yeo CJ, Salvia R, Buchler M, International Study Group on Pancreatic Surgery (ISGPS) (2017) The 2016 update of the International Study Group (ISGPS) definition and grading of postoperative pancreatic fistula: 11 Years After. *Surgery* 161(3):584–591. <https://doi.org/10.1016/j.surg.2016.11.014>
 27. Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I et al (2019) RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 366:l4898. <https://doi.org/10.1136/bmj.l4898>
 28. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M et al (2016) ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 355:i4919. <https://doi.org/10.1136/bmj.i4919>
 29. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J (2003) Methodological index for non-randomized studies (MINORS): development and validation of a new instrument. *ANZ J Surg* 73(9):712–716. <https://doi.org/10.1046/j.1445-2197.2003.02748.x>
 30. Downs SH, Black N (1998) 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* 52(6):377–384. <https://doi.org/10.1136/jech.52.6.377>
 31. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C (2013) Assessing bias in studies of prognostic factors. *Ann Intern Med* 158(4):280–286. <https://doi.org/10.7326/0003-4819-158-4-201302190-00009>
 32. Page MJ, McKenzie JE, Higgins JPT (2018) Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. *BMJ Open* 8(3):e019703. <https://doi.org/10.1136/bmjopen-2017-019703>
 33. Boutron I, Guittet L, Estellat C, Moher D, Hrobjartsson A, Ravaud P (2007) Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 4(2):e61. <https://doi.org/10.1371/journal.pmed.0040061>
 34. Probst P, Grummich K, Heger P, Zschke S, Knebel P, Ulrich A, Buehler MW, Diener MK (2016) Blinding in randomized controlled trials in general and abdominal surgery: protocol for a systematic review and empirical study. *Syst Rev* 5:48. <https://doi.org/10.1186/s13643-016-0226-4>
 35. Probst P, Ohmann S, Klaiber U, Huttner FJ, Billeter AT, Ulrich A et al (2017) Meta-analysis of immunonutrition in major abdominal surgery. *Br J Surg* 104(12):1594–1608. <https://doi.org/10.1002/bjs.10659>
 36. Corrigan N, Marshall H, Croft J, Copeland J, Jayne D, Brown J (2018) Exploring and adjusting for potential learning effects in ROLARR: a randomised controlled trial comparing robotic-assisted vs. standard laparoscopic surgery for rectal cancer resection. *Trials* 19(1):339. <https://doi.org/10.1186/s13063-018-2726-0>
 37. Wehrmann FS, de la Garza JR, Kowalewski KF, Schmidt MW, Muller K, Tapking C et al (2020) Learning curves of laparoscopic Roux-en-Y gastric bypass and sleeve gastrectomy in bariatric surgery: a systematic review and introduction of a standardization. *Obes Surg* 30(2):640–656. <https://doi.org/10.1007/s11695-019-04230-7>
 38. Nickel F, Haney CM, Muller-Stich BP, Hackert T (2020) Not yet IDEAL?—evidence and learning curves of minimally invasive pancreaticoduodenectomy. *Hepatobiliary Surg Nutr* 9(6):812–814. <https://doi.org/10.21037/hbsn.2020.03.22>
 39. Sterne JA, Gavaghan D, Egger M (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 53(11):1119–1129. [https://doi.org/10.1016/s0895-4356\(00\)00242-0](https://doi.org/10.1016/s0895-4356(00)00242-0)
 40. Begg, C. B., & Berlin, J. A. (1988). Publication bias - a problem in interpreting medical data. *Journal of the Royal Statistical Society Series a-Statistics in Society*, 151:419–463 <https://doi.org/10.2307/2982993>
 41. Sterne JAC, Harbord RM (2004) Funnel plots in meta-analysis. *Stata J* 4(2):127–141. <https://doi.org/10.1177/1536867x0400400204>
 42. Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50(4):1088–1101
 43. Egger M, Davey Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315(7109):629–634. <https://doi.org/10.1136/bmj.315.7109.629>
 44. Page, M. J., Higgins, J. P. T., & Sterne, J. A. C. (2021). Chapter 13.3.5.4 Tests for funnel plot asymmetry In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane handbook for systematic reviews of interventions* version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook.
 45. McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, Nicholl J (2009) No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 374(9695):1105–1112. [https://doi.org/10.1016/S0140-6736\(09\)61116-8](https://doi.org/10.1016/S0140-6736(09)61116-8)
 46. Blencowe NS, Boddy AP, Harris A, Hanna T, Whiting P, Cook JA, Blazeby JM (2015) Systematic review of intervention design and delivery in pragmatic and explanatory surgical randomized clinical trials. *Br J Surg* 102(9):1037–1047. <https://doi.org/10.1002/bjs.9808>
 47. Bender, R., & Lange, S. (2007). [The 2 by 2 table]. *Dtsch Med Wochenschr*, 132 Suppl 1, e12-14. doi:<https://doi.org/10.1055/s-2007-959029>
 48. Higgins JPT, Li T, Deeks JJ (editors) (2021) Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane handbook for systematic reviews of interventions* version 6.2 (updated February 2021). Available from www.training.cochrane.org/handbook.
 49. Borenstein M, Hedges L, Higgins J, Rothstein H (2009) *Introduction to meta-analysis*. West Sussex: John Wiley and Sons.
 50. Borenstein M, Hedges LV, Higgins JP, Rothstein HR (2010) *A basic introduction to fixed-effect and random-effects models for*

- meta-analysis. *Res Synth Methods* 1(2):97–111. <https://doi.org/10.1002/jrsm.12>
51. Ades AE, Lu G, Higgins JP (2005) The interpretation of random-effects meta-analysis in decision models. *Med Decis Mak* 25(6):646–654. <https://doi.org/10.1177/0272989X05282643>
 52. Brockwell SE, Gordon IR (2001) A comparison of statistical methods for meta-analysis. *Stat Med* 20(6):825–840. <https://doi.org/10.1002/sim.650>
 53. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Control Clin Trials* 7(3):177–188. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2)
 54. Normand SL (1999) Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med* 18(3):321–359. [https://doi.org/10.1002/\(sici\)1097-0258\(19990215\)18:3<321::aid-sim28>3.0.co;2-p](https://doi.org/10.1002/(sici)1097-0258(19990215)18:3<321::aid-sim28>3.0.co;2-p)
 55. Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. *Stat Med* 21(11):1539–1558. <https://doi.org/10.1002/sim.1186>
 56. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *BMJ* 327(7414):557–560. <https://doi.org/10.1136/bmj.327.7414.557>
 57. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ, GRADE Working Group (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336(7650):924–926. <https://doi.org/10.1136/bmj.39489.470347.AD>
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.