Langenbeck's Archives of Surgery (2018) 403:119–129
https://doi.org/10.1007/s00423-017-1646-x

ORIGINAL ARTICLE

Optimal literature search for systematic reviews in surgery

Käthe Goossen [1] · Solveig Tenckhoff [1] · Pascal Probst [1,2] · Kathrin Grummich [1] · André L. Mihaljevic [1,2] · Markus W. Büchler [2] · Markus K. Diener [1,2]

Received: 7 July 2017 / Accepted: 27 November 2017 / Published online: 5 December 2017

Abstract

Background The aim of the present study was to determine empirically which electronic databases contribute best to a literature search in surgical systematic reviews.
Methods For ten published systematic reviews, the systematic literature searches were repeated in the databases MEDLINE, Web of Science, CENTRAL, and EMBASE. On the basis of these reviews, a gold standard set of eligible articles was created. Recall (%), precision (%), unique contribution (%), and numbers needed to read (NNR) were calculated for each database, as well as for searches of citing references and of the reference lists of related systematic reviews (hand search).
Results CENTRAL yielded the highest recall (88.4%) and precision (8.3%) for randomized controlled trials (RCT), MEDLINE for non-randomized studies (NRS; recall 92.6%, precision 5.2%). The most effective combination of two databases plus hand searching for RCT was MEDLINE/CENTRAL (98.6% recall, NNR 97). Adding EMBASE marginally increased the recall to 99.3%, but with an NNR of 152. For NRS, the most effective combination was MEDLINE/Web of Science (99.5% recall, NNR 60).
Conclusions For surgical systematic reviews, the optimal literature search for RCT employs MEDLINE and CENTRAL. For surgical systematic reviews of NRS, Web of Science instead of CENTRAL should be searched. EMBASE does not contribute substantially to reviews with a surgical intervention.

Keywords Surgery · Systematic reviews · Literature search · Recall · Precision

Background

Systematic reviews (SR) are a key element of the clinical decision-making process. They are instrumental in transferring research results into evidence-based clinical practice and ensuring the quality of medical interventions. SR of randomized, controlled trials (RCT) represent the highest possible level of evidence on the *Oxford Levels of Evidence* scale [1]. Evidence-based surgery has become firmly rooted, with increasing numbers of RCT and SR published in the past two decades.

Ample methodological research has led to the definition of recommendations for the diverse facets of the SR process, compiled in the Cochrane Handbook and PRISMA statement and checklist [2, 3]. These sources do not, however, specify which databases systematic reviewers should choose beyond the general statement that "three bibliographic databases [are] generally considered to be the most important sources to search for reports of trials—CENTRAL, MEDLINE and EMBASE." The authors' personal experience in past systematic literature searches led to the hypothesis that EMBASE may not be the optimal database for surgical topics of interest.

Specific guidance for database selection can be found in the literature for SR in some medical areas [4–8] but not for surgical interventions. The selection is oriented only by general recommendations to search multiple classes of sources and all relevant sources within each class. These sources include bibliographic databases, reference lists, trials registries, and hand

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00423-017-1646-x) contains supplementary material, which is available to authorized users.

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

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

2 Department of General, Visceral and Transplantation Surgery, University Hospital Heidelberg, Im Neuenheimer Feld 110, 69120 Heidelberg, Germany

searching of journals [9, 10]. The absence of specific recommendations has led to a certain amount of uncertainty with respect to the number and choice of the most suitable bibliographic databases. For example, in the 17 SR published by the Study Center of the Germany Surgical Society (SDGC) in the past 5 years [11–27], multi-database searches were conducted in 90% of SR, but the selection of individual databases varied considerably. A median of three databases were searched, usually chosen among MEDLINE, CENTRAL, EMBASE, Web of Science/SCI, Biosis, CDSR, DARE, or LILACS. To maintain the highest quality standards while optimizing time and resource efficiency in the laborious SR process [28–31], it is desirable to minimize the number needed to read (NNR) [32]. A reduction in the number of irrelevant references retrieved by database searches may be achieved by reducing the number of databases that are accessed.

To date, it has not been shown empirically which database combination yields the optimum coverage of relevant references to studies for systematic reviews of surgical interventions. According to Sampson et al., the "selection of databases is ideally based on the potential contribution of each database to the project or on the potential for bias if a database is excluded, as supported by research evidence" [9].

The aim of the present study was to provide such evidence by empirically assessing the recall, precision, NNR, and unique contribution of the most important electronic databases for systematic reviews in the surgical literature. Furthermore, the contribution of alternate sources, i.e., screening of reference lists of preceding SR and of citing references (hand search), was to be evaluated. Based on these results, we intended to provide a first recommendation on which database combination to search when compiling SR of RCT and non-randomized studies (NRS) of surgical interventions, with the aim of maximizing coverage of relevant references, while minimizing non-relevant search results.

Methods

This empirical study was conducted according to a predefined protocol, provided as Additional file 1.

Generation of a representative set of systematic reviews

A representative set of systematic reviews reflecting current surgical research objectives was used as a basis to compile a gold standard set of references. These SR were identified by searching the three most high-impact journals of general surgery, i.e., *British Journal of Surgery, Annals of Surgery,* and *Journal of the American College of Surgeons* [33], for systematic reviews or meta-analyses published in 2012 in MEDLINE via Pubmed. This journal selection was arbitrary, based on an assumed greater likelihood of finding high-quality surgical SR in high-impact surgical journals.

The search strategy was: (("The British journal of surgery" [jo]) OR ("Annals of surgery" [jo]) OR ("Journal of the American College of Surgeons" [jo])) AND ((meta-analysis [tiab]) OR (systematic review [tiab])) AND ("2012/01/01" [pdat]: "2012/12/31" [pdat]).

Ninety references published in the three top journals of surgery in 2012 were identified by the search strategy. A set of systematic reviews was selected based on the following criteria:

- Systematic review that follows the PRISMA statement and has searched both MEDLINE and EMBASE, as recommended in the Cochrane Handbook
- Coverage of a research question in an area of general and abdominal surgery
- Inclusion of at least 15 RCT/NRS
- Intervention review in the fields of surgical strategy, medical device, perioperative drug, perioperative nutrition or other surgical research questions
- Non-overlapping with other included systematic reviews in terms of surgical intervention and included studies. In case of overlap, the review with the largest number of included studies was used.
- Systematic reviews of diagnostic studies and of studies without surgical intervention were excluded.

The flow of literature references is presented in Fig. 1. Sixty-seven references were excluded following screening the titles and abstracts, and a further four based on the full text. The reasons for exclusion were that they described no surgical intervention (n = 19), were no systematic review (n = 16), included < 15 studies (n = 12), has not searched both MEDLINE and EMBASE (n = 11), reviewed a diagnostic test (n = 7), or were in an area other than general or abdominal surgery (n = 6). Details are documented in Additional file 2. The 19 eligible systematic reviews were sorted by Medline accession number. Following exclusion of two additional SR (n = 1 subject overlap, i.e., surgical site infection, n = 1 topic too vast to repeat literature search, i.e., predictors of surgical performance), the ten most recent SR were retained. They covered a wide spectrum of relevant research questions (see Additional file 3) [34–43].

From the included systematic reviews, the literature search strategies or key search terms were extracted, as well as inclusion/exclusion criteria.

Pilot searches and selection of the most relevant databases

Two pilot searches based on the topics of two systematic reviews were performed without time or language restrictions in

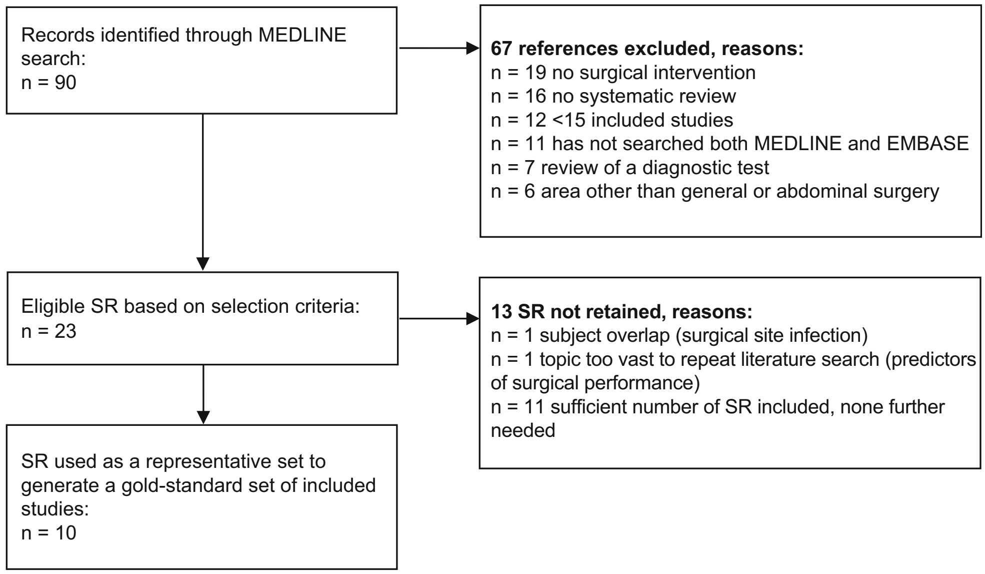

Fig. 1 Study flow in the search and selection of SR

the electronic databases listed below. To attain sound results, these searches were selected to cover research questions in surgery, one rather common topic from a systematic review that included only RCT, as well as one more unusual, less well-researched topic from a systematic review that encompassed also NRS. Thus, O'Reilly et al. [38] investigated laparoscopic versus open repair of primary unilateral inguinal hernia, with limits to RCT published in English, and Zeng et al. [42] studied laparoscopic versus open distal gastrectomy for early gastric cancer in a primarily Asian population with a majority of NRS. The following electronic databases were searched in the pilot phase: MEDLINE, EMBASE, Cochrane Library (covering CENTRAL, CDSR, and DARE), Web of Science (also covering Biosis), PsycInfo, CINAHL, and LILACS. Google Scholar was not searched, based on recent evidence that necessary elements for systematic scientific literature retrieval are not available [44], and because of excessive screening times [45].

Because of the consistently low numbers of references to surgically relevant studies identified in CINAHL ($n = 3$ and 2, respectively), PsycInfo ($n = 0$), and LILACS ($n = 0$), all further investigations were limited to the four remaining databases, i.e., MEDLINE, Web of Science, CENTRAL, and EMBASE.

Generation of a gold standard set of references

The key search terms or search strategies extracted from the ten selected SR were used as a basis to develop our own search algorithms adapted to each database. Four databases were searched without time or language restrictions for each of the ten selected SR topics with these algorithms. RCT filters were employed only for SR topics for which the authors had excluded NRS [37, 43]. The search results were transferred to Endnote (version X7.5) and labeled with the database of origin. Irrelevant references were excluded following screening of titles and abstracts, and the full texts were retrieved for potentially eligible references. In addition to electronic databases, references cited in related systematic reviews were screened. At the end of the literature screening process, citing references (defined as articles that have included a reference in their bibliographies that was previously identified as eligible) were retrieved from the Science Citation Index via Web of Science and reviewed analogously.

The inclusion/exclusion criteria from the original systematic reviews were applied to the identified trials by one reviewer to determine eligibility and checked by the other. Any disagreements were resolved by consensus. All eligible references, complemented by any references included in the published reviews but not identified by our search, were retained as a gold-standard set of references.

Analysis of database performance

The set of references found in each database was assessed against this gold standard. Venn diagrams were used to present the number of references identified through each of the four included databases.

For each database, the following analytical measures were computed using a customized file in Microsoft Excel 2010, each with 95% confidence intervals assuming normal distribution of the effects:

- % recall (sensitivity) as weighted means with 95% confidence intervals (number of relevant reports identified through this source (true positives) divided by total number of relevant reports in existence (true positives + false negatives))

- % precision as weighted means with 95% confidence intervals (number of relevant reports identified through this source (true positives) divided by total number of reports identified through this source (true positives + false positives))
- Number needed to read (NNR) with 95% confidence intervals (1/precision)
- % contribution of "single-database references" (number of reports uniquely contributed by this database divided by total number of relevant reports in existence (true positives + false negatives)) as weighted means with 95% confidence intervals
- Retrieval odds ratios computed in analogy to diagnostic odds ratios [46] and corresponding 95% confidence intervals, representing the association of relevant reports identified through each source with the total number of relevant reports within the full sample (true positives/false positives)/(false negatives/true negatives)

A contingency table is presented in the supplemental information (see Additional file 4).

The results were stratified by type of study (RCT, NRS).

For "single-database references," i.e., the number of reports uniquely contributed by one database, the question whether these could also be identified by hand search was also examined.

Finally, the precision and recall of combinations of the best databases were computed.

Results

Generation of a gold standard set of references

The ten selected SR covered a range of topics in surgery. The median number of references screened per single SR was 694. Taken together, the SR reviewed a sum of 265 references to relevant studies. A median of 25 studies (RCT and/or NRS) and 1890 patients were included per SR (Table 1). Typically, the electronic databases searched by the review authors included MEDLINE, EMBASE, and CENTRAL, often complemented with one other database.

Systematic literature searches for each of their topics were repeated in MEDLINE, Web of Science, CENTRAL, and EMBASE. Fifty-one thousand four hundred eighty-five were retrieved by electronic database searching, or a median of 4155 per SR. The records identified were screened based on the inclusion/exclusion criteria defined in each SR. This yielded a set of 509 included publications, complemented by 16 publications identified by hand searching (cited references search in Web of Science and citation lists of related reviews) and two not found by our search but cited in the original SR.

The gold standard set thus encompassed 527 references, of which 265 (50.3%) were cited in the original SR. One hundred fifty-nine references (30.2%) were published subsequently. One hundred one references (19.2%) were published within the search period of the SR but not identified and cited by the SR authors, because our search was likely more exhaustive than that of the SR authors in terms of search terms employed and databases accessed (see Additional file 5). For RCT, the proportion cited in the original SR was larger (63.3%), with 25 RCT (17.0%) missed by the review authors and 28 RCT (19.0%) published subsequently.

Analysis of database performance

Searches in three large literature databases MEDLINE, Web of Science, and EMBASE each retrieved a large proportion of the gold standard set of 527 references to relevant studies (481, 444, and 461, respectively, Fig. 2a), whereas the recall in CENTRAL was lower (146 references). The other sources that were investigated, i.e., the reference lists of preceding SR and citation searching, recalled 399 references. The unique contribution by literature databases was largest, in absolute terms, for Web of Science ($n = 13$ references uniquely identified by this database, of which 4 were, moreover, not covered by the reference lists of preceding SR or by citation searching) and MEDLINE ($n = 11$, of which 7 were not covered by the reference lists of preceding SR or by citation searching). The reference lists of preceding SR or by citation searching contributed 16 additional unique references not found in any database.

For the 147 RCT in the gold standard set, searches in all four investigated databases led to comparable levels of recall (129 for MEDLINE, 119 for Web of Science, 130 for CENTRAL, and 128 for EMBASE, Fig. 2b). All databases except EMBASE contributed one unique RCT reference to the gold standard set. The most important contributors of NRS references were MEDLINE (352/380), EMBASE (333/380), and Web of Science (325/380). Web of Science and MEDLINE identified most unique references (12 and 10, respectively).

Table 2 gives an overview of the quantitative results obtained, stratified by references to randomized vs. non-randomized studies.

The results for electronic databases showed that MEDLINE had the highest recall overall and for NRS, closely followed by EMBASE and Web of Science. CENTRAL is a database focused on RCT but ca. 20% of indexed studies are high-quality non-randomized, controlled clinical trials. Unsurprisingly, it was the most sensitive in retrieving RCT, but added little when searching for NRS. The precision was highest for CENTRAL, the database with the lowest number of references retrieved, and lowest for EMBASE.

Table 1 Characteristics of included SR

SR	Topic	Databases searched	Refs. screened (*n*)	Studies (*n*)	Patients (*n*)
Bhangu [34]	Outcomes following emergency surgery for *Clostridium difficile* infections	MEDLINE, EMBASE, CENTRAL	461	31 NRS	1433
Constantinides [35]	Retroperitoneoscopic vs. laparoscopic adrenalectomy	MEDLINE, EMBASE, CENTRAL Ovid[a]	682	2 RCT, 20 NRS	1966
Fung [36]	Safety and complication profile of colonic SILS	MEDLINE, EMBASE, CENTRAL	71	38 NRS	565
Marimuthu [37]	Immune modulating nutrition vs. standard diet for open GI surgery	MEDLINE, EMBASE, CENTRAL SCI	1214	26 RCT	2496
O'Reilly [38]	Laparoscopic vs. open repair of primary unilateral inguinal hernia	MEDLINE, EMBASE	2280	27 RCT (34 refs)	7161
Shabanzadeh [39]	Laparoscopic vs. open surgery decreases surgical site infection in obese patients	MEDLINE, EMBASE, CENTRAL	2981	8 RCT, 36 NRS	59,370
Trastulli [43]	Single-incision vs. conventional laparoscopic cholecystectomy	MEDLINE, EMBASE, CENTRAL, CINAHL	241	13 RCT	923
Venkat [40]	Laparoscopic vs. open distal pancreatectomy	MEDLINE, EMBASE, CENTRAL	818	18 NRS	1814
Xiong [41]	Stent vs. no stent in pancreaticoduodenectomy	MEDLINE, EMBASE, CENTRAL	705	5 RCT, 11 NRS	1726
Zeng [42]	Laparoscopic vs. open gastrectomy for early gastric cancer	MEDLINE, EMBASE, CENTRAL, CKRID	827	5 RCT, 18 NRS	3411

[a] It remains unclear from the publication which database was searched via Ovid, which is not in itself a bibliographic database but a search platform

BJS, British Journal of Surgery; CKRID, China Knowledge Resource Integrated Database; CWG, Cochrane Wounds Group; ERIC, Educational Resources Information Centre; SILS, single-incision laparoscopic surgery; WoS, Web of Science

The unique contribution rate was greatest for MEDLINE and Web of Science, both for RCT and NRS, as well as CENTRAL for RCT. The % contribution of "single-database references" were substantially lower for EMBASE. CENTRAL did not contribute unique NRS. It is important to note that no RCT filter was used in 8/10 searches within the present study, so that when searching specifically for RCT and using a filter, the NNR will effectively be lower.

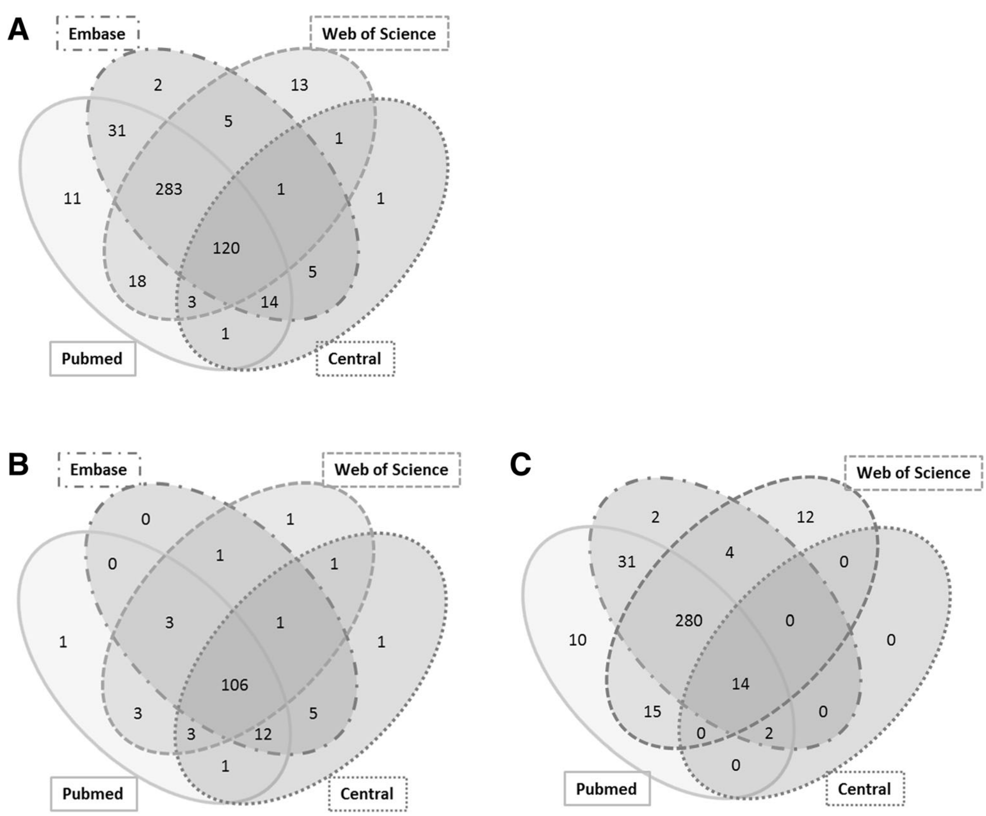

Fig. 2 Venn diagrams with number of studies identified by database (MEDLINE via Pubmed): **a** for all studies, **b** for RCT only, and **c** for NRS only

Table 2 Recall, precision, NNR, unique contribution, and retrieval odds ratio by electronic database, stratified by RCT and NRS

	Recall (%)	Precision (%)	NNR (*n*)	Unique contribution (%)	Retrieval OR
RCT					
MEDLINE	87.8 [82.2–93.3]	1.9 [1.4–2.4]	53 [42–73]	0.7 [0.0–1.4]	16.0 [10.7–21.3]
WoS	81.0 [75.3–86.6]	1.5 [1.2–1.7]	68 [57–85]	0.7 [0.0–1.5]	7.0 [3.2–10.8]
CENTRAL	88.4 [83.6–93.2]	8.3 [5.2–11.4]	12 [9–19]	0.7 [0.0–1.5]	136.8 [28.7–245.0]
EMBASE	87.1 [82.7–91.4]	1.7 [0.9–2.5]	58 [40–105]	0	6.2 [2.4–9.9]
Hand search	68.7 [48.4–89.0]	2.1 [1.4–2.9]	47 [34–73]	4.8 [2.2–7.4]	6.4 [−1.2–14.1]
NRS					
MEDLINE	92.6 [89.1–96.2]	5.2 [4.4–6.0]	19 [17–23]	2.6 [0.5–4.7]	28.5 [−8.1–65.1]
WoS	85.5 [77.3–93.8]	3.9 [2.7–5.0]	26 [20–36]	3.2 [1.4–4.9]	9.8 [0.9–18.8]
CENTRAL	4.2 [0–9.2]	3.9 [1.0–6.8]	26 [15–97]	0	0.7 [−7.3–8.8]
EMBASE	87.6 [84.2–91.0]	2.7 [2.1–3.4]	37 [30–47]	0.5 [0.0–1.1]	6.5 [3.7–9.4]
Hand search	78.4 [66.7–90.2]	4.8 [3.4–6.1]	21 [16–29]	2.4 [0.2–4.6]	10.8 [−2.3–23.9]

Data are presented as weighted means with 95% CI in square brackets

The retrieval odds ratio, which describes the odds of retrieving a reference by searching a specific database, allowed ranking the four databases by search efficiency. For RCT, by far the most efficient database was CENTRAL, followed by MEDLINE, Web of Science, and EMBASE. For NRS, the ranking was found to be MEDLINE, then Web of Science, EMBASE, and CENTRAL.

There were 27 "single-database references" in sum, 11 from MEDLINE, 13 from Web of Science, 1 from CENTRAL, and 2 from EMBASE (see Fig. 2a; Additional file 6). For these, an assessment was also made to determine whether they could alternatively be identified by other means that should be part of any systematic review search strategy, such as the screening of citation lists or citing references.

Table 3 Recall, precision, and NNR for database combinations, stratified by study type

	Recall (%)	Precision (%)	NNR (*n*)
RCT			
ME/WoS/EM/CE/HS	99.3 [98.5–100]	0.5 [0.4–0.6]	212 [169–285]
ME/WoS/CE/HS	99.3 [98.5–100]	0.7 [0.5–0.8]	152 [127–191]
ME/EM/CE/HS[a]	99.3 [98.5–100][a]	0.7 [0.5–0.9][a]	152 [116–219][a]
ME/WoS/HS	95.9 [92.9–98.9]	0.7 [0.6–0.9]	141 [116–179]
ME/WoS	90.5 [85.9–95.0]	0.9 [0.7–1.0]	115 [96–145]
ME/CE/HS[b]	98.6 [97.8–99.5][b]	1.0 [0.8–1.3][b]	97 [79–125][b]
ME/CE	92.5 [89.8–95.2]	1.4 [1.1–1.7]	70 [58–87]
ME/EM/HS	98.6 [97.1–100]	0.8 [0.5–1.0]	132 [98–202]
ME/EM	92.5 [89.2–95.9]	0.9 [0.6–1.3]	107 [78–170]
NRS			
ME/WoS/EM/CE/HS	99.7 [99.3–100]	1.1 [0.9–1.3]	94 [78–116]
ME/WoS/CE/HS	99.5 [98.7–100]	1.6 [1.3–2.0]	61 [51–75]
ME/EM/CE/HS[a]	98.7 [96.6–100][a]	1.4 [1.1–1.7][a]	71 [60–87][a]
ME/WoS/HS[b]	99.5 [98.7–100][b]	1.7 [1.4–2.0][b]	60 [51–73][b]
ME/WoS	96.8 [93.9–99.8]	2.4 [1.9–2.8]	42 [35–52]
ME/CE/HS	98.2 [95.9–100]	2.7 [2.2–3.1]	38 [32–45]
ME/CE	92.6 [89.1–96.2]	4.9 [4.1–5.7]	20 [17–24]
ME/EM/HS	98.7 [96.6–100]	1.4 [1.2–1.7]	70 [59–85]
ME/EM	94.2 [91.1–97.3]	1.9 [1.5–2.2]	53 [45–66]

CE, CENTRAL; EM, EMBASE; HS, hand search; ME, MEDLINE accessed via Pubmed; WoS, Web of Science

[a] Standard combination

[b] Recommended combination

This showed that MEDLINE and Web of Science contributed most references that were not identified by hand searching (7 and 4, respectively). Moreover, the results highlight the importance of searching citing references, which contributed 14 unique references and found 14 out of the 27 single-database references. SR citation lists contributed a further 2 unique references and detected 5 out of the 27 single-database references. Overall, hand searching had a combined recall of 68.7%, precision of 2.1%, retrieval OR of 6.4, and unique contribution rate of 4.8% for RCT, and a recall of 78.4%, precision of 4.8%, retrieval OR of 10.8, and unique contribution rate of 2.4% for NRS.

Precision and recall of database combinations

Table 3 illustrates the recall, precision, and NNR obtained for several database combinations, stratified by RCT and NRS.

The complete set of databases (MEDLINE, Web of Science, CENTRAL, EMBASE) and hand searching was found to reach > 99% recall for both RCT and NRS, but with a large NNR of close to 100 for NRS and over 200 for RCT.

For RCT, the typically used combination of MEDLINE, CENTRAL, EMBASE, and hand searching was computed to have a recall of 99.3% and precision of 0.7%, corresponding to an NNR of 152. In comparison, a combination of the two most effective databases MEDLINE and CENTRAL plus citation searching and checking the reference lists of relevant papers accounted for a recall of 98.6% of the gold standard set of citations. The overall precision for this combination was 1.0%, resulting in an NNR of 97. Thus, omitting EMBASE led to a 36% reduction of the NNR with a marginal loss in recall (0.7%). The contribution of hand searching to this database combination was substantial (6.1% greater recall). For NRS, the highest recall from two databases plus hand searching was obtained when searching MEDLINE and Web of Science, with a combined recall of 99.5%, precision of 1.7%, and NNR of 60. In comparison, the standard search in MEDLINE, Central, EMBASE, and hand searching led to a recall of 98.7%, precision of 1.4%, and NNR of 71. A search in MEDLINE, Web of Science, and hand searching thus leads to a marginally improved recall by 0.8% and reduction in NNR by 16%.

In contrast, the original SR, which also searched a median of 3 databases, had a recall of 77.0% [95% CI, 64.6%–89.5%] of RCT and 71.4% [95% CI, 61.7%–81.1%] of NRS based on the gold standard set of references published within the relevant time period.

Discussion

To provide an evidence-based recommendation on which database combination to search when compiling a SR in the area of surgery, systematic reviews with a mean of 25 included studies were investigated. For SR of this size, the results have made clear that the location of relevant references is a time- and resource-intensive research activity, involving screening of thousands of potentially eligible references retrieved by database searches. For this reason, any gain in efficiency is welcome. Other researchers have explored, for example, the use of automated tools for the generation of the search strategy [47], screening of references [48–51], as well as techniques for rapid reviews [52, 53].

The goal of our empirical study was to reduce the numbers needed to read by defining the most relevant set of electronic databases to search when compiling surgical SR. Computation of a retrieval odds ratio for the most relevant databases allowed ranking their utility in the following order: MEDLINE, Web of Science, CENTRAL, and EMBASE for NRS and CENTRAL, MEDLINE, Web of Science, and EMBASE for RCT.

MEDLINE is generally recognized as the first database to search for any topic in medicine. It is large, well-indexed, and free of charge. When accessed via Pubmed, it is the most up-to-date database. The great utility of CENTRAL for locating RCT is well known and constitutes the primary aim of this database. More surprising was that among the other three large databases, EMBASE ranked third for each study type. For RCT, this may at least in part be due to the re-publication of EMBASE records of RCTs in CENTRAL starting 1996 [54]. This is mirrored by the strong overlap between EMBASE and CENTRAL for the RCT retrieved (124 retrieved by both out of a total of 131 retrieved by either, or 95%). Another factor was certainly the large absolute number of references retrieved, leading to a larger proportion of irrelevant records. Moreover, according to several authors, EMBASE has a particular emphasis on the pharmaceutical literature [55, 56]. Web of Science compared favorably with EMBASE in our study, even though its search functions are less sophisticated and no subject index is available. A likely reason is the coverage of different journals, and the less refined search tools may actually lead to more sensitive searches.

For RCT, the computed NNR were larger than expected, presumably because no RCT filter was used in most searches, which targeted both randomized and non-randomized studies. Therefore, in real-life situations, the NNR can be expected to be brought down substantially by the use of RCT filters.

Because researchers are typically under substantial time and resource constraints, a combination search of MEDLINE and CENTRAL plus hand searching may be considered sufficient for RCT (98.6% recall, 1.0% precision, NNR of 97). For NRS, the highest recall from two databases plus hand searching was obtained when searching MEDLINE and Web of Science, with a combined recall of 99.5%, precision of 1.7% and NNR of 60. These precision values are in line with those reported for a large survey of published SR, i.e., 2.9% (1.3–8.1%) [57].

Even these limited searches have yielded substantially better recall than the original SR. In the context of the variable individual search efficiency even for SR in top journals [58], the quality of the search strategy is a highly relevant contributor to the completeness of the evidence. During the review process of any systematic review, journals should verify that, as recommended by AMSTAR criteria, (a) at least two electronic sources were searched, (b) the report includes years and databases used, (c) key words and/or MESH terms are stated, (d) the search strategy is provided, and (e) searches were supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found [59]. A useful resource is "An evidence-based practice guideline for the peer review of electronic search strategies" by Sampson et al. [60].

Another contributor is the use of citation searching. Starting from a relevant reference, it allows searching forward in time and identifying other articles that have included this particular reference in their bibliographies, so that the method is complementary to keyword-based literature searching. Citation searching is provided as an electronic function by certain literature database providers such as Web of Science and Scopus. The importance of citation searching has previously been assessed empirically with a more limited dataset in an unrelated medical field [61]. Wright et al. concluded based on the low number of unique references that they retrieved that this approach may not make the best use of time and resources. In contrast, in our hands, searching the citing references for each of the 527 gold standard references via Web of Science retrieved a larger number of unique references beyond the four databases searched (n = 7) than either EMBASE (n = 2) or CENTRAL (n = 1) did beyond the other three.

Instead of adding another major database such as EMBASE to the recommended search, systematic review authors may consider searching regional databases from countries where most research in the field originates [6]. An example would be East Asian databases for clinical studies on gastric cancer, as was done in the included SR by Zeng et al. [42].

Discrepancies from the study as planned

According to the protocol, stratification of the results by publication year was planned. The results were not included in the final manuscript, because the authors considered they did not add anything novel to the research question but are available as supplemental files (see Additional files 7 and 8). In short, stratification by publication year showed that (a) the citation lists of other SR were a good source of earlier references but did not contribute the most recent ones; (b) citing references contributed an over-proportional amount of citations from the most recent 5 years; and (c) no major gap in indexing became apparent from our data.

In the protocol, we intended to use a figure to quantify the value of adding a database to the search of all others, the "NNR for one additional unique reference." This would describe the numbers of additional articles that would need to be screened in order to obtain one further relevant reference from the database in question and was calculated as the NNR divided by the % contribution of references unique to a single database. However, the main message obtained from these data, namely that EMBASE contributed least effectively to the search results, is conveyed both more clearly and more effectively by the results in Table 3. Therefore, these results were left out of the manuscript.

Strengths and limitations

This is the first comprehensive empirical study quantifying the utility of electronic literature databases in surgery. Previous related work, performed in other medical areas, was more limited in scope, and analytical measures reported were typically restricted to recall and precision [5, 62, 63].

The results have led to a clear recommendation as to which databases to search for surgical SR, with the potential of improving cost- and resource-effectiveness without substantial losses in recall. Our results create confidence that methodologically sound SR may be performed even without access to the costly database EMBASE [64].

The original choice of journals to identify relevant SR may have influenced the computed recall of the published SR searches, if one assumes that the journal impact factor correlates with the quality of the literature search. This does not, however, affect the principal message of the present study.

Due to a large number of records that needed to be screened, searches and screening were not performed in duplicate, and the bibliographies of the 527 included references were not hand searched. As a result, some records may have been missed. However, the present study duplicated the searches of published SR, thus effectively performing the third survey of the literature on each research question, with numerous additional included references. In addition, no bias for one database over another is to be expected because the authors were unaware of the origin of each record at the time of screening.

A major limitation of the study is the heavy reliance on large bibliographic databases in the process of compiling the gold standard set of references, which may have introduced database bias [65]. However, this was complemented by the references cited in published SR, of which 2/527 (0.4%) had not been found by our searches, as well as by SR citation lists, which contributed two additional references (0.4%). In view of these small numbers, the resulting potential bias will be minimal.

The literature search and screening was performed by expert systematic reviewers, so that retrieval may differ in practice depending on the level of search expertise [66]. Other authors have also detected deficiencies in recall among published systematic reviews attributable to the search strategy, with 22/522 gold standard references not retrieved from any database using the reported searches [63]. Our fraction of 95 references not detected out of a gold standard of 360 within the search period of the SR (26%) was still greater.

As discussed above, no RCT filter was applied to most research questions, with the exception of two [37, 43]. For SR based purely on RCT, the effectiveness of searching Web of Science over EMBASE, as quantified by the retrieval odds ratio as well as the recall, may be lower because of the lack of validated RCT filters for Web of Science [67, 68].

Finally, the results of the present study are limited to surgical SR, and the validity for other medical fields remains untested.

Conclusions

For SR in the field of surgical interventions, the databases MEDLINE and CENTRAL need to be searched as a minimum when the study type is limited to RCT, and supplementation with a third database has limited benefit. For SR including NRS, Web of Science should be added. CENTRAL may be omitted in cases where no RCT are expected. Screening of citation lists and citing references contributes substantially to a comprehensive literature search. EMBASE plays a lesser role for surgical SR.

Authors' contributions MD and KGo conceived the work. KGo, KGr, PP, and MD designed the protocol. KGo, ST, and KGr acquired the data. KGo and ST performed the analysis. KGo, ST, PP, AM, MB, and MD interpreted the results. KGo and PP drafted the manuscript. ST, KGr, AM, MB, and MD revised it critically. All authors approved the final version of the manuscript.**Funding** The resources and facilities of the University of Heidelberg were used in conducting this study. There was no external source of funding.

Compliance with ethical standards

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and material The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests The authors declare that they have no competing interests.

Sources of financial support The resources and facilities of the University of Heidelberg were used in conducting this study. There was no external source of funding.

Abbreviations CE, CENTRAL; EM, EMBASE; HS, Hand search; ME, MEDLINE accessed via Pubmed; NNR , Numbers needed to read; NRS, Non-randomized study/studies; RCT, Randomized controlled trial(s); SR , Systematic review(s); WoS, Web of Science

References

1. OCEBM Levels of Evidence Working Group*: "The Oxford 2011 levels of evidence". **Oxford Centre for Evidence-Based Medicine (*OCEBM Levels of Evidence Working Group*=Jeremy Howick ICJLL, Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, Bob Phillips HT, Olive Goddard and Mary Hodgkinson)**. http://www.cebm.net/index.aspx?o=5653. Accessed 5 Jul 2016
2. McKenzie S, Mailey B, Artinyan A, Kim J, Ellenhorn JD (2010) The incidence and outcomes of pancreatectomy in patients with metastatic pancreatic adenocarcinoma. JOP: Journal of the Pancreas [Electric Resource] 11(4):341–347
3. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097
4. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F (1994) Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 120(8):667–676. https://doi.org/10.7326/0003-4819-120-8-199404150-00008
5. Preston L, Carroll C, Gardois P, Paisley S, Kaltenthaler E (2015) Improving search efficiency for systematic reviews of diagnostic test accuracy: an exploratory study to assess the viability of limiting to MEDLINE, EMBASE and reference checking. Syst Rev 4(1):82. https://doi.org/10.1186/s13643-015-0074-7
6. Betran AP, Say L, Gulmezoglu AM, Allen T, Hampson L (2005) Effectiveness of different databases in identifying studies for systematic reviews: experience from the WHO systematic review of maternal morbidity and mortality. BMC Med Res Methodol 5(1):6. https://doi.org/10.1186/1471-2288-5-6
7. Ogilvie D, Hamilton V, Egan M, Petticrew M (2005) Systematic reviews of health effects of social interventions: 1. Finding the evidence: how far should you go? J Epidemiol Community Health 59(9):804–808. https://doi.org/10.1136/jech.2005.034181
8. Royle P, Bain L, Waugh N (2005) Systematic reviews of epidemiology in diabetes: finding the evidence. BMC Med Res Methodol 5(1):2. https://doi.org/10.1186/1471-2288-5-2
9. Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R, St John PD, Viola R, Raina P (2003) Should meta-analysts search Embase in addition to Medline? J Clin Epidemiol 56(10):943–955. https://doi.org/10.1016/S0895-4356(03)00110-0
10. Vassar M, Carr B, Kash-Holley M, DeWitt E, Koller C, Day J, Day K, Herrmann D, Holzmann M (2015) Database choices in endocrine systematic reviews. J Med Libr Assoc 103(4):189–192. https://doi.org/10.3163/1536-5050.103.4.005
11. Contin P, Goossen K, Grummich K, Jensen K, Schmitz-Winnenthal H, Buchler MW et al (2013) ENERgized vessel sealing systems versus CONventional hemostasis techniques in thyroid surgery—the ENERCON systematic review and network meta-analysis. Langenbeck's Arch Surg 398(8):1039–1056. https://doi.org/10.1007/s00423-013-1137-7
12. Harnoss JC, Ulrich AB, Harnoss JM, Diener MK, Buchler MW, Welsch T (2014) Use and results of consensus definitions in pancreatic surgery: a systematic review. Surgery 155(1):47–57. https://doi.org/10.1016/j.surg.2013.05.035
13. Harnoss JM, Harnoss JC, Diener MK, Contin P, Ulrich AB, Buchler MW et al (2014) Portal annular pancreas: a systematic review of a clinical challenge. Pancreas 43(7):981–986. https://doi.org/10.1097/MPA.0000000000000186
14. Mehrabi A, Fischer L, Hafezi M, Dirlewanger A, Grenacher L, Diener MK, Fonouni H, Golriz M, Garoussi C, Fard N, Rahbari NN, Werner J, Büchler MW (2014) A systematic review of localization, surgical treatment options, and outcome of insulinoma.

Pancreas 43(5):675–686. https://doi.org/10.1097/MPA.0000000000000110
15. Huttner FJ, Koessler-Ebs J, Hackert T, Ulrich A, Buchler MW, Diener MK (2015) Meta-analysis of surgical outcome after enucleation versus standard resection for pancreatic neoplasms. Br J Surg 102(9):1026–1036. https://doi.org/10.1002/bjs.9819
16. Huttner FJ, Tenckhoff S, Jensen K, Uhlmann L, Kulu Y, Buchler MW et al (2015) Meta-analysis of reconstruction techniques after low anterior resection for rectal cancer. Br J Surg 102(7):735–745. https://doi.org/10.1002/bjs.9782
17. Klaiber U, Probst P, Knebel P, Contin P, Diener MK, Buchler MW et al (2015) Meta-analysis of complication rates for single-loop versus dual-loop (Roux-en-Y) with isolated pancreaticojejunostomy reconstruction after pancreaticoduodenectomy. Br J Surg 102(4):331–340. https://doi.org/10.1002/bjs.9703
18. Loffler T, Rossion I, Goossen K, Saure D, Weitz J, Ulrich A et al (2015) Hand suture versus stapler for closure of loop ileostomy—a systematic review and meta-analysis of randomized controlled trials. Langenbeck's Arch Surg 400(2):193–205. https://doi.org/10.1007/s00423-014-1265-8
19. Mehrabi A, Hafezi M, Arvin J, Esmaeilzadeh M, Garoussi C, Emami G, Kössler-Ebs J, Müller-Stich BP, Büchler MW, Hackert T, Diener MK (2015) A systematic review and meta-analysis of laparoscopic versus open distal pancreatectomy for benign and malignant lesions of the pancreas: it's time to randomize. Surgery 157(1):45–55. https://doi.org/10.1016/j.surg.2014.06.081
20. Muller-Stich BP, Kenngott HG, Gondan M, Stock C, Linke GR, Fritz F et al (2015) Use of mesh in laparoscopic paraesophageal hernia repair: a meta-analysis and risk-benefit analysis. PLoS One 10(10):e0139547. https://doi.org/10.1371/journal.pone.0139547
21. Muller-Stich BP, Senft JD, Warschkow R, Kenngott HG, Billeter AT, Vit G et al (2015) Surgical versus medical treatment of type 2 diabetes mellitus in nonseverely obese patients: a systematic review and meta-analysis. Ann Surg 261(3):421–429. https://doi.org/10.1097/SLA.0000000000001014
22. Probst P, Huttner FJ, Klaiber U, Knebel P, Ulrich A, Buchler MW et al (2015) Stapler versus scalpel resection followed by hand-sewn closure of the pancreatic remnant for distal pancreatectomy. Cochrane Database Syst Rev 11:Cd008688
23. Huttner FJ, Fitzmaurice C, Schwarzer G, Seiler CM, Antes G, Buchler MW et al (2016) Pylorus-preserving pancreaticoduodenectomy (pp Whipple) versus pancreaticoduodenectomy (classic Whipple) for surgical treatment of periampullary and pancreatic carcinoma, Cd006053. Cochrane Database Syst Rev 2
24. Huttner FJ, Mihaljevic AL, Hackert T, Ulrich A, Buchler MW, Diener MK (2016) Effectiveness of Tachosil((R)) in the prevention of postoperative pancreatic fistula after distal pancreatectomy: a systematic review and meta-analysis. Langenbeck’s Arch Surg 401(2):151–159. https://doi.org/10.1007/s00423-016-1382-7
25. Kossler-Ebs JB, Grummich K, Jensen K, Huttner FJ, Muller-Stich B, Seiler CM et al (2016) Incisional hernia rates after laparoscopic or open abdominal surgery—a systematic review and meta-analysis. World J Surg 40(10):2319–2330. https://doi.org/10.1007/s00268-016-3520-3
26. 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
27. 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
28. Fraser C, Murray A, Burr J (2006) Identifying observational studies of surgical interventions in MEDLINE and EMBASE. BMC Med Res Methodol 6(1):41. https://doi.org/10.1186/1471-2288-6-41
29. Beyer FR, Wright K (2013) Can we prioritise which databases to search? A case study using a systematic review of frozen shoulder management. Health Inf Libr J 30(1):49–58. https://doi.org/10.1111/hir.12009
30. Booth A (2010) How much searching is enough? Comprehensive versus optimal retrieval for technology assessments. Int J Technol Assess Health Care 26(4):431–435. https://doi.org/10.1017/S0266462310000966
31. Egger M, Juni P, Bartlett C, Holenstein F, Sterne J (2003) How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 7(1):1–76
32. Bachmann LM, Estermann P, Kronenberg C, ter Riet G (2003) Identifying diagnostic accuracy studies in EMBASE. J Med Libr Assoc 91(3):341–346
33. Journal Citation Reports® (Thomson Reuters, 2016)
34. Bhangu A, Nepogodiev D, Gupta A, Torrance A, Singh P (2012) Systematic review and meta-analysis of outcomes following emergency surgery for Clostridium difficile colitis. Br J Surg 99(11):1501–1513. https://doi.org/10.1002/bjs.8868
35. Constantinides VA, Christakis I, Touska P, Palazzo FF (2012) Systematic review and meta-analysis of retroperitoneoscopic versus laparoscopic adrenalectomy. Br J Surg 99(12):1639–1648. https://doi.org/10.1002/bjs.8921
36. Fung AK, Aly EH (2012) Systematic review of single-incision laparoscopic colonic surgery. Br J Surg 99(10):1353–1364. https://doi.org/10.1002/bjs.8834
37. Marimuthu K, Varadhan KK, Ljungqvist O, Lobo DN (2012) A meta-analysis of the effect of combinations of immune modulating nutrients on outcome in patients undergoing major open gastrointestinal surgery. Ann Surg 255(6):1060–1068. https://doi.org/10.1097/SLA.0b013e318252edf8
38. O’Reilly EA, Burke JP, O'Connell PR (2012) A meta-analysis of surgical morbidity and recurrence after laparoscopic and open repair of primary unilateral inguinal hernia. Ann Surg 255(5):846–853. https://doi.org/10.1097/SLA.0b013e31824e96cf
39. Shabanzadeh DM, Sorensen LT (2012) Laparoscopic surgery compared with open surgery decreases surgical site infection in obese patients: a systematic review and meta-analysis. Ann Surg 256(6):934–945. https://doi.org/10.1097/SLA.0b013e318269a46b
40. Venkat R, Edil BH, Schulick RD, Lidor AO, Makary MA, Wolfgang CL (2012) Laparoscopic distal pancreatectomy is associated with significantly less overall morbidity compared to the open technique: a systematic review and meta-analysis. Ann Surg 255(6):1048–1059. https://doi.org/10.1097/SLA.0b013e318251ee09
41. Xiong JJ, Altaf K, Mukherjee R, Huang W, WM H, Li A et al (2012) Systematic review and meta-analysis of outcomes after intraoperative pancreatic duct stent placement during pancreaticoduodenectomy. Br J Surg 99(8):1050–1061. https://doi.org/10.1002/bjs.8788
42. Zeng YK, Yang ZL, Peng JS, Lin HS, Cai L (2012) Laparoscopy-assisted versus open distal gastrectomy for early gastric cancer: evidence from randomized and nonrandomized clinical trials. Ann Surg 256(1):39–52. https://doi.org/10.1097/SLA.0b013e3182583e2e
43. Trastulli S, Cirocchi R, Desiderio J, Guarino S, Santoro A, Parisi A, Noya G, Boselli C (2013) Systematic review and meta-analysis of randomized clinical trials comparing single-incision versus conventional laparoscopic cholecystectomy. Br J Surg 100(2):191–208. https://doi.org/10.1002/bjs.8937
44. Boeker M, Vach W, Motschall E (2013) Google scholar as replacement for systematic literature searches: good relative recall and precision are not enough. BMC Med Res Methodol 13(1). https://doi.org/10.1186/1471-2288-13-131

45. Levay P, Ainsworth N, Kettle R, Morgan A (2016) Identifying evidence for public health guidance: a comparison of citation searching with Web of Science and Google Scholar. Research Synthesis Methods 7(1):34–45. https://doi.org/10.1002/jrsm.1158
46. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM (2003) The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 56(11):1129–1135. https://doi.org/10.1016/S0895-4356(03)00177-X
47. Hausner E, Waffenschmidt S, Kaiser T, Simon M (2012) Routine development of objectively derived search strategies. Syst Rev 1(1): 19. https://doi.org/10.1186/2046-4053-1-19
48. Rathbone J, Hoffmann T, Glasziou P (2015) Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers. Syst Rev 4(1):80. https://doi.org/10.1186/s13643-015-0067-6
49. Shemilt I, Simon A, Hollands GJ, Marteau TM, Ogilvie D, O'Mara-Eves A et al: Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews. Res Synth Methods 2014, 5(1):31–49, DOI: https://doi.org/10.1002/jrsm.1093
50. Wallace BC, Trikalinos TA, Lau J, Brodley C, Schmid CH (2010) Semi-automated screening of biomedical citations for systematic reviews. BMC Bioinformatics 11(1):55. https://doi.org/10.1186/1471-2105-11-55
51. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S (2015) Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev 4(1):5. https://doi.org/10.1186/2046-4053-4-5
52. Polisena J, Garritty C, Umscheid CA, Kamel C, Samra K, Smith J, Vosilla A (2015) Rapid review summit: an overview and initiation of a research agenda. Syst Rev 4:111. https://doi.org/10.1186/s13643-015-0111-6
53. Tricco AC, Antony J, Zarin W, Strifler L, Ghassemi M, Ivory J, Perrier L, Hutton B, Moher D, Straus SE (2015) A scoping review of rapid review methods. BMC Med 13(1):224. https://doi.org/10.1186/s12916-015-0465-6
54. Lefebvre C, Eisinga A, McDonald S, Paul N (2008) Enhancing access to reports of randomized trials published world-wide—the contribution of EMBASE records to the Cochrane Central Register of Controlled Trials (CENTRAL) in the Cochrane Library. Emerg Themes Epidemiol 5(1):13. https://doi.org/10.1186/1742-7622-5-13
55. Hallam E, Plaice C (1999) An evaluation of EMBASE within the NHS: findings of the Database Access Project working partnership to extend the knowledge base of healthcare. Health Libr Rev 16(3): 192–203. https://doi.org/10.1046/j.1365-2532.1999.0225a-1.x
56. Robinson KA (2005) Should meta-analysts search Embase in addition to Medline? J Clin Epidemiol 58(3):320; **author reply 1**. https://doi.org/10.1016/j.jclinepi.2004.08.001
57. Sampson M, Tetzlaff J, Urquhart C (2011) Precision of healthcare systematic review searches in a cross-sectional sample. Research Synthesis Methods 2(2):119–125. https://doi.org/10.1002/jrsm.42
58. Sampson M, McGowan J (2011) Inquisitio validus index Medicus: a simple method of validating MEDLINE systematic review searches. Research Synthesis Methods 2(2):103–109. https://doi.org/10.1002/jrsm.40
59. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM (2007) Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 7(1):10. https://doi.org/10.1186/1471-2288-7-10
60. Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C (2009) An evidence-based practice guideline for the peer review of electronic search strategies. J Clin Epidemiol 62(9):944–952. https://doi.org/10.1016/j.jclinepi.2008.10.012
61. Wright K, Golder S, Rodriguez-Lopez R (2014) Citation searching: a systematic review case study of multiple risk behaviour interventions. BMC Med Res Methodol 14(1):73. https://doi.org/10.1186/1471-2288-14-73
62. Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B (2016) The contribution of databases to the results of systematic reviews: a cross-sectional study. BMC Med Res Methodol 16(1):127. https://doi.org/10.1186/s12874-016-0232-1
63. Whiting P, Westwood M, Burke M, Sterne J, Glanville J (2008) Systematic reviews of test accuracy should search a range of databases to identify primary studies. J Clin Epidemiol 61(4):357–364. https://doi.org/10.1016/j.jclinepi.2007.05.013
64. Reveiz L, Ospina E, Zorrilla AF (2004) Should we consider Embase in Latin America? J Clin Epidemiol 57(8):866; **author reply 7-8**–866; author reply 868. https://doi.org/10.1016/j.jclinepi.2004.02.005
65. Egger M, Smith GD (1998) Bias in location and selection of studies. BMJ 316(7124):61–66. https://doi.org/10.1136/bmj.316.7124.61
66. McKibbon KA, Haynes RB, Dilks CJ, Ramsden MF, Ryan NC, Baker L et al (1990) How good are clinical MEDLINE searches? A comparative study of clinical end-user and librarian searches. Comput Biomed Res 23(6):583–593. https://doi.org/10.1016/0010-4809(90)90042-B
67. For an up-to-date list of search filters, see: The InterTASC Information Specialists' Sub-Group Search Filter Resource, Search Filters by Study Design. https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/filters-to-identify-randomized-controlled-trials-and. Accessed 4 Jul 2016. Scottish Intercollegiate Guidelines Network (SIGN) search filters, http://www.sign.ac.uk/methodology/filters.html. Accessed 4 Jul 2016
68. For a non-validated Web of Science RCT filter, see: University of Alberta Libraries, Systematic Reviews-Searching the Literature, Randomized Controlled Trial Filters, Web of Science. http://guides.library.ualberta.ca/c.php?g=248586&p=1655962. Accessed 4 Jul 2016