Technological Forecasting & Social Change 212 (2025) 123833

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

How to search for literature in systematic reviews and meta-analyses: A comprehensive step-by-step guide

Michael Gusenbauer [a,*], Sebastian P. Gauster [b]

[a] Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz, Austria
[b] University of Innsbruck, Universitätsstraße 15, 6020 Innsbruck, Austria

ARTICLE INFO

Keywords:
Literature sampling
Literature search methods
Database choice
Methodological guidance
Systematic literature review
Meta-analysis

ABSTRACT

We introduce a comprehensive guide to improve literature sampling in systematic literature reviews (SLRs) and meta-analyses (MAs) in management and adjacent social science disciplines. Analyzing 404 SLRs and MAs from top-tier management-related journals and the 135 methodological studies on which they relied revealed significant opportunities to improve sampling quality. We found that inadequate guidance and an overreliance on unfounded "best practices" often leads to suboptimal search strategies and incomplete reporting. To improve literature sampling, we have developed a guide that details four sequenced and prioritized scoping, searching, screening, and reporting steps. The guide also includes recommendations for the most suitable literature databases to search with, suggested (AI-enabled) tool support, and two checks to validate the effectiveness of search strategies. We created this guide by adapting the state-of-the-art recommendations of the Cochrane Handbook and PRISMA—the widely acknowledged benchmarks in methodical and reporting guidance—and supplementing them with additional evidence-based methodological insights. By adopting the guide, authors can ensure their sampling is effective and meets the high standards of SLRs and MAs, thereby likely boosting the impact of their review. The guide is applicable to all review types that require comprehensive, transparent, and reproducible sampling, including bibliometric studies and integrative reviews.

1. Introduction

Systematic literature reviews (SLRs) and meta-analyses (MAs) are crucial for developing management theory (Post et al., 2020) and making management knowledge accessible to practice (Briner et al., 2009). They reduce the fragmentation of management knowledge (Chen and Hitt, 2021), ease the *"burden of knowledge"* (Jones, 2009, p. 283), and help researchers and practitioners stay abreast of skyrocketing publication numbers (Gusenbauer, 2021).

SLRs and MAs can only fulfill these functions well when they rigorously and comprehensively review all relevant literature in a transparent and reproducible fashion (Hiebl, 2023; Tranfield et al., 2003). The sampling process determines the quality of the literature analyzed (Wanyama et al., 2022), the extent of database bias (Egger and Smith, 1998), the scope for drawing representative conclusions, and the statistical power of meta-analytical models (Valentine et al., 2010). Only thorough sampling ensures that the many hours spent reviewing literature yield high-quality reviews that can truly be termed *"systematic"* (Simsek et al., 2021). Indeed, better sampling seems to consistently relate to higher citation counts in top management journals (Simsek et al., 2021).

Unfortunately, many SLRs and MAs fall short of the high quality standards their readers expect (Ioannidis, 2016; Moore et al., 2022), especially in the field of management (Gusenbauer, 2021). Our analysis of 404 SLRs and MAs published in high-quality management-related journals reveals significant room for improvement in methodological rigor in literature sampling. Reviews often suffer from narrow search scopes, suboptimal database use, inadequate search strings, and a lack of transparent reporting, leading to significant omissions of relevant documents, inferior representativeness, inefficiencies, and non-replicability. These issues permeate journals of every level of quality and affect authors, peer reviewers, and editors alike.

The crux of the problem is the inadequacy of current guidance for conducting literature sampling (Kunisch et al., 2018, p. 519). Our investigation into the guidelines used by management reviews to inform their sampling revealed a contradiction: While these guides are often invaluable in areas other than sampling, they are not sufficient to guide authors on the intricate decisions that make sampling sufficiently

* Corresponding author.
E-mail addresses: michael.gusenbauer@jku.at (M. Gusenbauer), sebastian.gauster@uibk.ac.at (S.P. Gauster).

https://doi.org/10.1016/j.techfore.2024.123833
Received 27 February 2024; Received in revised form 12 August 2024; Accepted 12 October 2024
Available online 16 December 2024

systematic. We found that the guides that authors relied on were often outdated, not evidence-based, not directly applicable to management reviews, and sometimes even downright misleading (see analysis in Supplementary Materials 6–8). While many authors clearly make earnest efforts to base their sampling on recommendations, these are often unfounded "best practices" that merely propagate inadequate sampling practices.

In response to these challenges, our research seeks to assist authors, editors, and reviewers of SLRs and MAs by offering practical, well-founded sampling recommendations in the form of a step-by-step guide that encompasses the entire process of scoping, search, screening, and reporting. In doing so, we aim to tackle two critical yet unresolved questions:

(1) What is the state of the art in systematic literature sampling methods, and how can it be tailored to management SLRs and MAs?
(2) How do top management SLRs and MAs sample literature, and to what extent does their sampling follow state-of-the-art methods?

Methodologically, our guide rests on two pillars. First, our recommendations to management scholars producing SLRs and MAs are based on the most extensively scrutinized sampling guidance available—the state of the art in literature sampling methods. Similar to Tranfield et al. (2003) and many others in management and economics (e.g., Kugley et al., 2017; Kunisch et al., 2023; Ringquist, 2013; Stanley and Doucouliagos, 2012), we adapt Cochrane Handbook guidance (Higgins et al., 2023), which is *"widely regarded as a benchmark [in SLR methodology]"* (Briner et al., 2009, p. 25) and PRISMA as the standard in reporting. Furthermore, we draw on the TARCiS statement (Hirt et al., 2024) for supplementary related guidance on citation-searching and rely on other methodological reviews to adapt this guidance to management. As databases become increasingly differentiated in terms of coverage and functionalities, searching the right ones is ever more important (Hiebl, 2023; Wilczynski et al., 2004). Based on novel scientometric insights into the databases that management researchers currently use (plus some capable alternatives that they should be aware of), the paper recommends particularly suitable database combinations for better sample identification. We synthesize, sequence, and prioritize these recommendations in a four-step guide to SLR sampling.

Second, to learn about current sampling practices, we conduct a methodological review (Aguinis et al., 2023) of the search, screening, and reporting procedures of 404 SLRs and MAs. To understand the origin of sampling practices, we also review the 135 methodological studies that guided the sampling strategies of these 404 reviews. To learn about areas for improvement, we compare these contemporary sampling practices in management SLRs and MAs to state-of-the art sampling methods.

The four-step guide contributes to improving sampling quality in three main ways: First, the guide's evidence-based recommendations are intended to replace the unfounded sampling recommendations on which most management reviews currently rely. To draw prescriptive methodical guidance, the guide makes the wealth of insights in the Cochrane Handbook and other expert methodological guidance accessible to researchers beyond the health sciences. Second, the four-step process clearly sequences sampling steps to give authors procedural clarity. The guide meaningfully integrates iterative, exploratory steps essential to accounting for social sciences' diverse concepts and language (Boell and Cecez-Kecmanovic, 2015), clear sequencing to make the process highly comprehensible, and prioritization to account for researchers' individual situations (some steps are mandatory, highly desirable, or optional). The guide includes two quality checks to help validate the diligence of sampling strategies and also suggests specific procedural steps whose efficiency may be enhanced by (semi-) automated and AI-assisted tools. Based on the guide's recommended minimum standards, evaluative sampling questions, and quality checks, authors, peer reviewers, and editors should be able to contribute to better sampling outcomes while freeing up time for other essential aspects of the review. Third, the guide offers a comprehensive overview of the prevailing methodological practices in management SLRs and MAs. By assessing current management practices, this overview not only identifies gaps and potential areas for advancement in literature sampling but also establishes a baseline against which future improvements can be compared.

Overall, these sampling improvements should translate into higher citation impact (Simsek et al., 2021) of SLRs, MAs, and other review methods that aim for systematicity in literature sampling. The guide is relevant for management, yet also beyond, as the sampling logics are similar in adjacent social science disciplines such as psychology, economics, or decision sciences that share similar difficulties in capturing literatures (Papaioannou et al., 2010).

2. Background

2.1. The scope of this guide

Review guidance frequently cited in management, and using a logic similar to that of the Cochrane Handbook (Sharma and Bansal, 2020; Tranfield et al., 2003; Xiao and Watson, 2019), broadly distinguishes three stages review authors should pass through: 1) planning the review; 2) conducting the review; and 3) reporting and dissemination. This sampling guide mainly focuses on the phases of *searching the literature* and *screening for inclusion* as part of the "conducting the review" stage. These two phases also touch on the preceding *developing and validating the review protocol* phase via thorough and iterative *scoping* detailing the coverage and purpose of the review (Xiao and Watson, 2019). Accordingly, searching and screening are connected to an iterative learning process that determines the review's scope to ensure it is both large enough to include a sufficient number of relevant documents and small enough to permit meaningful review depth. The sampling process is succeeded by a *quality assessment* of the identified studies (a succinct description of the subsequent steps can be found in, e.g., Xiao and Watson, 2019). The scope of the sampling process is outlined in Fig. 1: Overview of the four-step literature sampling process.

Conducting comprehensive, transparent, and reproducible searching and screening is *"one of the fundamental differences between a traditional narrative review and a systematic review"* (Tranfield et al., 2003, p. 215). Comprehensiveness is *"judged by its contribution to answering the review's […] predefined research question(s)"* through a representative synthesis of the literature defined by set inclusion and exclusion criteria (Hiebl, 2023). The quality of the review will *"highly depend on the literature collected"* (Kunisch et al., 2023; Xiao and Watson, 2019, p. 103). Transparency ensures the independent validation of the review process and allows reproducibility of the review by others. While many literature review methods exist, this methodical guidance purposefully focuses on *systematic* literature reviews and meta-analyses, whose sampling adheres to particularly high standards.

Many MAs implicitly aim to synthesize all relevant research, as they often seek to include as many studies as possible to achieve a large sample size (Stanley and Doucouliagos, 2012). Borenstein et al. (2009, pp. xxvii—xxviii) note: *"A meta-analysis can be applied to any data, but if the goal of the analysis is to provide a synthesis of a body of data from various sources, then it is usually imperative that the data be compiled as part of a systematic review… If the search strategy is flawed in concept or execution, or if the studies are providing biased results, then problems exist in the review that the meta-analysis cannot correct."* Therefore, this guidance applies both to MAs and to any other type of review, such as bibliometric studies (Zupic and Čater, 2015) or integrative reviews (Cronin and George, 2023; Rousseau, 2024), that aims to rigorously synthesize all research on a topic.

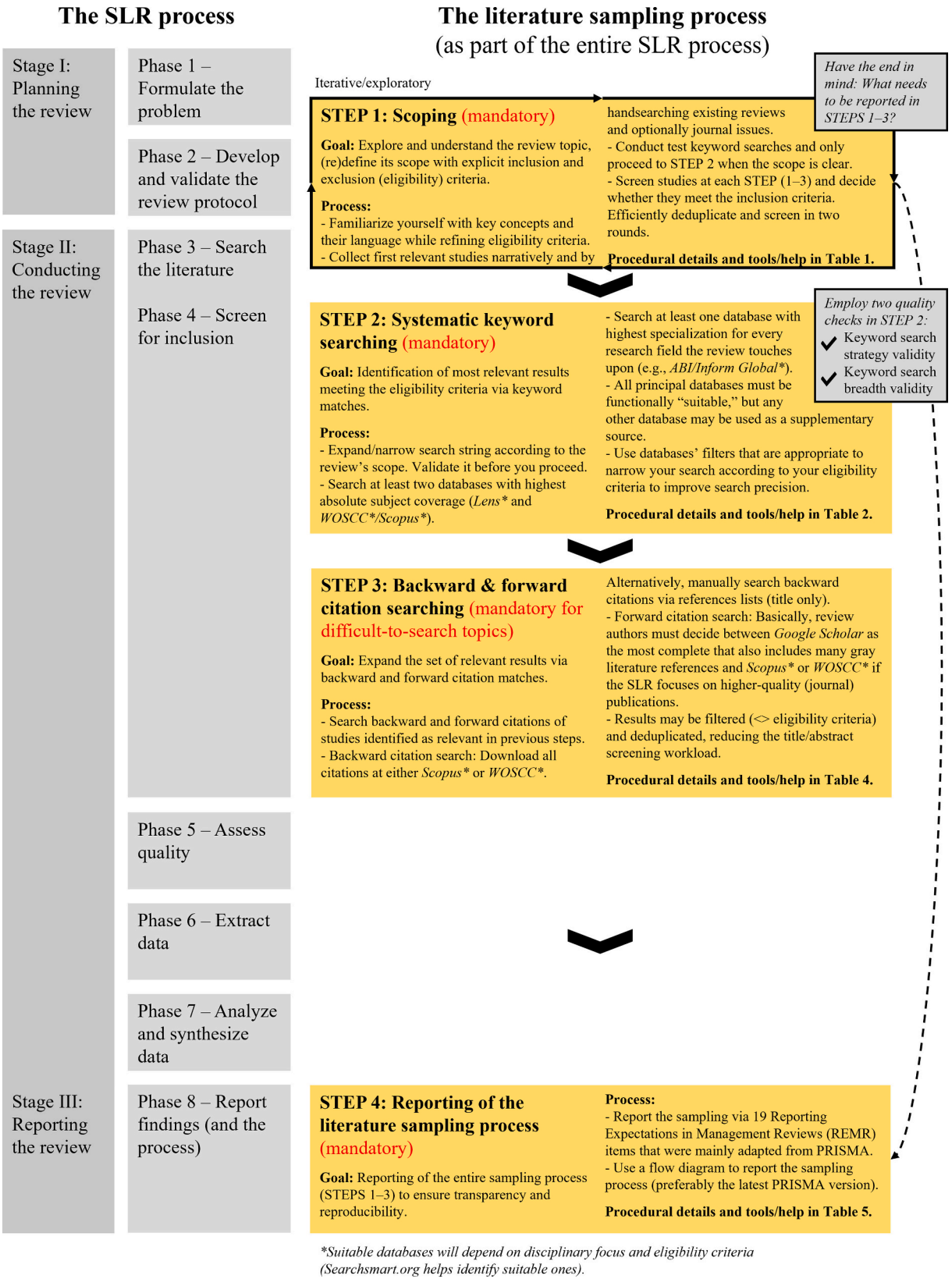

Fig. 1. Overview of the four-step literature sampling process.

2.2. Extensive resource requirements of systematic reviewing work

Systematic reviews are frequently seen as the gold standard of review types (Munn et al., 2018) and *"have become regarded as the most reliable form of research review"* (Denyer and Tranfield, 2006, p. 217). Nevertheless, their unmatched levels of rigor and comprehensiveness make systematic reviews labor-intensive, requiring up to 24 months of dedicated review work, a quarter of which can be attributed to the sampling

phase (Collins et al., 2015; Dicks et al., 2014; Haddaway and Westgate, 2019; McGowan and Sampson, 2005). Thus, before starting a SLR, authors should assess whether they have the necessary resources to conduct the review.

Due to these substantial resource requirements, alternative review methods such as *rapid reviews* (Garritty et al., 2021), *rapid evidence assessments* (Barends and Rousseau, 2018), or *tailored approaches* (Cooper et al., 2022; Cooper et al., 2018a) have emerged to loosen the comprehensiveness criterion in sampling. Further, if resources are limited, authors may want to consider less exhaustive or less rigorous types of review (see, e.g., Grant and Booth (2009), Paré et al. (2015), or Gruner and Minunno (2023)). Once authors have determined the type of review they wish to pursue, they must also follow the expected sampling requirements. The recommendations in this sampling guide should help authors manage their review burden by improving sampling effectiveness while adhering to these high sampling standards.

2.3. Adapting medical guidance to management

In the early 2000s, when SLRs became popular in management (Gusenbauer, 2021), the SLR methodology was adopted from medicine. Particularly, the seminal article by Tranfield et al. (2003), which translated concepts from medicine to management, became the guiding framework for thousands of management SLRs. Tranfield et al. (2003, pp. 213–214) described management's systematic review methods of the time as being in their *"infancy,"* portraying review sampling methods as *"implicitly idiosyncratic," and "biased"* with methodological choices *"not recorded"* and precise inclusion/exclusion criteria *"often not formally agreed, applied, recorded, or monitored."*

To advance management's literature review methodology, Tranfield and co-authors went on to broadly recommend Cochrane's methodology for systematic reviews and qualified specific decision points to management (e.g., Tranfield et al., 2003). Thus, management SLRs borrowed from the medicinal conduct guidance, which had been *"[…] developed out of a need to ensure that decisions affecting people's lives can be informed by an up-to-date and complete understanding of the relevant research evidence"* (Higgins et al., 2020, p. 3). Many other fields have since adopted the Cochrane Handbook (Higgins et al., 2023), which has become *the* standard systematic review guidance across many disciplines as it is *"relevant to all authors of systematic reviews"* (Cumpston et al., 2019, pp. 1—2).

2.3.1. Differences between medicine and management

When applying the Cochrane Handbook's recommendations, it is important to consider sampling-related differences between medicine and management. First, while medical SLRs focus on empirical evidence of interventions (Higgins et al., 2023), largely disregarding theory (Greenhalgh, 2021), management SLRs synthesize a broader range of literature types. In terms of *knowledge synthesis*, SLRs may acquire all sorts of management literature, both empirical and logic-based, in both a positivist and an interpretivist fashion, and both management-related and interdisciplinary, to tackle knowledge fragmentation (Chen and Hitt, 2021). In terms of *evidence-based management*, SLRs are used *"for locating, appraising, synthesizing, and reporting 'best evidence'"* to support *"managerial decisions"* on focused research questions (Briner et al., 2009, pp. 21, 24).

Second, while medicine has fairly standardized language for diagnoses, diseases, or treatments, concepts in social science (and thus management) are less clearly defined, and names are often ambiguous (Boell and Cecez-Kecmanovic, 2015). As a consequence for sampling, searches will be less precise, making their design even more important, and supplementary search techniques such as citation searching and scoping of research topics will be disproportionately more important in management (and other social sciences) compared to medicine.

To account for these differences between medical and management SLRs, this guide translates the authoritative recommendations of the Cochrane Handbook, also by consulting additional methodological sources (particularly PRISMA, TARCiS, and Search Smart) that help qualify specific methodological recommendations and suggest relevant bibliographic databases.

3. Method

We use two analyses to answer our research questions and develop the four-step sampling guide. The first derives sampling recommendations by reviewing and structuring established guidance. The second analyzes current sampling strategies in management SLRs and MAs and compares them to the four-step guide.

3.1. Developing sampling guidance for management

To derive sampling recommendations for management SLRs and MAs, we adapt conduct and reporting recommendations separately. This separation is strongly emphasized as essential in medicine (Caulley et al., 2020; Logullo et al., 2020; Sarkis-Onofre et al., 2021).

3.1.1. Conduct: developing methodological expectations of management reviews (MEMR)

Medical SLR guidance has already informed methodological studies in management (Durach et al., 2017; Tranfield et al., 2003). As this translated guidance is not detailed enough to cover the many considerations of good sampling, we chose to review and adapt medical guidance at the level of single methodological recommendations. We used the Cochrane Handbook's authoritative *Methodological Expectations of Cochrane Intervention Reviews (MECIR)* items (Higgins et al., 2023) and the *TARCiS statement* (Hirt et al., 2023) as the basis for our sampling guidance. MECIR stipulates items (or procedures) for *searching for studies (C24–C38)* and *selecting studies to include in the review (C39–C42)*. The *TARCiS statement* (Hirt et al., 2024)—a Delphi study of an international panel of methodological experts—lists 10 specific recommendations on citation searching.

To qualify specific conduct items, we used additional methodological studies—particularly those we identified as most relevant for management's sampling questions (Adams et al., 2017; Hiebl, 2023; Tranfield et al., 2003). To rate the necessity of specific methodological sampling recommendations, we adopted MECIR's system. The application of each of the items can be either *"mandatory"* or *"highly desirable,"* where *"items described as mandatory are expected to be applied, and if they are not, then an appropriate justification should be provided… Items described as highly desirable should generally be implemented, but there are reasonable exceptions and justifications are not required"* (Higgins et al., 2020, p. 7). This allowed us to tailor the MECIR and TARCiS items to management (details see Supplementary Material 3). We called the adapted items *"Methodological Expectations of Management Reviews (MEMR)"* and included them in the corresponding sections of our four-step guide (see Tables 1–2 and 4–5).

To develop sampling recommendations tailored to management, we also needed to derive database recommendations. To recommend the most productive databases for sampling management documents, we assessed the coverage and functionality of academic databases and the search systems through which they can be accessed. We used the original methodology (Gusenbauer, 2019, 2022, 2024a; Gusenbauer and Haddaway, 2020, 2021) of Search Smart (Gusenbauer, 2024b) a free website that compares academic databases, and extended the available databases to include those frequently relied on by management scholars. We collected information on all databases that were used in at least 1 %

(i.e., four or more times) of the 404 SLRs and MAs we analyzed and the databases listed by Hiebl (2023). We added the information on these databases and search systems[1] to Search Smart, which allowed us to compare 101 database options to identify the most productive for keyword searches and backward and forward citation searches in management SLRs and MAs. This comparison can easily be replicated for other disciplines.

3.1.2. Reporting: developing reporting expectations in management reviews (REMR)

To derive solid reporting guidance for management, we adapted the most established reporting guidance available: the PRISMA statement (Page et al., 2021) and its search extension PRISMA-S (Rethlefsen et al., 2021). PRISMA is endorsed by Cochrane (Higgins et al., 2023) and is most widely used and endorsed, including in the social sciences (Havránek et al., 2020). PRISMA is a multidisciplinary reporting guide (Leclercq et al., 2019) that presents an *"evidence-based minimum set of items for reporting in systematic reviews and meta-analyses"* (Moher et al., 2009, p. 128). To make the guide's reporting recommendations as complete as possible, we also included reporting advice from the TARCiS statement on citation searching (Hirt et al., 2024) and reporting items were created specifically for management-related disciplines: the Revised Meta-Analysis Reporting Standards (MARS) (Appelbaum et al., 2018) for meta-analyses in psychology, QMARS (Levitt et al., 2018) for qualitative meta-analyses in psychology, and the MAER guidelines (Havránek et al., 2020) for meta-analyses in economics. As not all of the PRISMA items were relevant for management (e.g., reporting of register searches), and many were overlapping, we synthesized them to derive a list of "reporting expectations in management reviews" (REMRs) suggested for management SLRs and MAs (for details, see Supplementary Material 4).

3.2. Analyzing the status quo of literature searching in SLRs and MAs in management

3.2.1. Data collection

We want to learn about the current literature search practices in journals relevant to management researchers, particularly those rated higher quality. Accordingly, our methodological review (Aguinis et al., 2023) included a cross-sectional analysis of literature search strategies across all 3-, 4-, and 4*-rated journal publications according to the 2021 AJG ranking. The AJG ranking covers *"a broad set of journals in which business and management academics may seek to publish their research"* (Chartered Association of Business Schools, 2021, p. 6). We opted for a cross-sectional focus on the status quo (as opposed to historical perspectives; i.e., Hiebl (2023)) to compare current sampling practice with what the best methodical guidance currently recommends for high-quality sampling.

In composing our search string, we prioritized precision over recall (for details on our sampling, see Supplementary Material 2) to draw a sample that was representative without being exhaustive. We searched title fields only, as most systematic reviews and meta-analyses are identified as such in their titles. On 7 March 2022, we identified 435 SLRs and MAs published in management journals in 2021. We included both SLRs and MAs in our analysis because these methodologies equally require rigorous sampling to synthesize a broad set of studies effectively, as emphasized by various methodological guidance (e.g., Borenstein et al., 2009; Higgins et al., 2023; Steel et al., 2021). We downloaded all 435 studies and organized them in citation software, after which both authors independently screened first the studies' titles/abstracts and then their full texts for relevance. In total, we excluded 31 book reviews and methodological articles that did not meet our inclusion criteria, leading to a final sample ($n = 404$) consisting of 199 SLRs (49 %), 171 MAs (42 %), and 34 (8 %) studies that were both MAs and SLRs (for a full list, see Supplementary Material 9).

3.2.2. Analysis

We derived our coding items to assess searching and screening practices from the PRISMA 2020 checklist items (Page et al., 2021) *(5) eligibility criteria, (6) information sources, (7) search strategy, (8) selection,* and *(17) study characteristics;* from the PRISMA-S extension (Rethlefsen et al., 2021), and from a review of databases' functionality assessments (Gusenbauer and Haddaway, 2020). Both authors jointly coded the 404 SLRs and MAs following the structure of 60 coding items, for a total of 24,240 individual codings.

We used the following codings: First, seven items on study characteristics and retrieved information sources included article information on the systematic procedure, search time, and applied data sources (e.g., databases, registers, websites, and reference lists). Second, 47 items are associated with the reviews' search strategy (e.g., keyword search, forward citation, backward citation), the Boolean operators (e.g., "AND," "OR," "NOT"), search limiters (e.g., journal quality, study types, timeframe), and search results (e.g., the number of studies identified, screened, or included). Third, three items are linked to reporting standards and supplementary material. Hence, we coded whether review authors followed flowcharts (e.g., PRISMA) and included additional resources (e.g., full list of articles included). Fourth, we included three items to capture whether the methodological section of each review reported guidance by conduct guidance, database guidance, or best practice guidance for the sampling process. All 135 identified conduct and database guidance items—including newly published conduct guidance—were reviewed in terms of selected sampling recommendations to rate their adequacy and identify areas for improvement. The results of the analysis of guidance used to steer sampling in management SLRs and MAs are described and detailed in Supplementary Materials 6–8.

3.2.3. Comparison of the four-step guide with current sampling procedures in management

We compare the state-of-the-art practices of the four-step guide with current sampling procedures in management. Since each study has a different scope, making detailed analyses of search strategies difficult, our analysis focuses on the quantitative comparison of SLR and MA authors' choices. Thus, we discuss the prevalence of selected scoping, searching, screening, and reporting choices that were related to the four-step guide's recommendations. To illustrate differences in sampling quality and provide learning opportunities, we describe cases that employed both exemplary and instructively suboptimal procedures.

4. A comprehensive step-by-step guide to systematic literature sampling

In this section, we describe a well-founded guide for sample identification that meets the three central principles of SLRs (and MAs): comprehensiveness (identifying all studies within the eligibility criteria), transparency, and reproducibility. Following established guidance in medicine (Higgins et al., 2023; Carol Lefebvre et al., 2019; Xiao and Watson, 2019), we recommend a search strategy that builds on keyword searching (Cooper et al., 2018a; Gusenbauer and Haddaway, 2020), enhanced by a combination of supplementary search heuristics whose importance is determined by the review topic's search difficulty.

The recommended workflow most productively sequences four sampling steps. All high-quality management SLRs must start with a

[1] We use the same terminology as in Gusenbauer and Haddaway (2020) to distinguish between search systems and databases: *"Search system: any resource that allows the user to search for academic records on the web, including any type of search engine, search platform or bibliographic database that might be used for this purpose. The user interacts with the search interface of a given search system to execute searches* via *queries, browsing and filtering. Database: The underlying dataset of citation records accessed* via *a search engine, a platform or any other search system."*

thorough scoping phase (STEP 1: handsearching, expert contacts, narrative searching including previous reviews), building the foundation with iterative and exploratory steps that lead to learning and useful preliminary results. The well-scoped topic is then meaningfully expanded by language-based (STEP 2: keyword searching) and associative search steps (STEP 3: backward and forward citation searching). Finally, the entire sampling process and its results need to be thoroughly documented to ensure transparency and reproducibility (STEP 4). The sequence and relevance-rating of these steps should improve the effectiveness of reviews in identifying more relevant documents while bolstering author efficiency by limiting the workload involved in search and screening work.

We discuss critical considerations and background explanations for each step in this section and detail the procedural descriptions, decision points, help, and (AI-assisted) tool suggestions in Tables 1–2 and 4–5. The tables also include a column with evaluation questions that review authors and peer reviewers can use to assess a review's sampling quality (informed by Barends et al. (2017); Hiebl (2023); Simsek et al. (2021); Tranfield et al. (2003). Database recommendations are detailed in Supplementary Material 1. Boxes 1-4 following each of the four steps describes how management review authors currently apply the step in question.

4.1. STEP 1: Scoping the review (for details, see Table 1)

Prior to the actual systematic search process that identifies most relevant studies, authors must specify the scope of their review (Higgins et al., 2023) to *"assess the relevance and size of the literature and to delimit the subject area or topic"* (Tranfield et al., 2003, p. 214). The scope of the review is at the discretion of the authors (Hiebl, 2023). Once scope has been determined by the research question and meaningful inclusion and exclusion criteria (also called "eligibility criteria"), SLR authors should aim to identify all the literature that falls within the scope of the review in a transparent and reproducible manner.

To decide on the scope of their review, authors must deeply and iteratively reflect on their envisioned research topic, its central concepts, and research questions of interest (Simsek et al., 2024). Extensive scoping of the review topic means refining the research question and inclusion/exclusion criteria to isolate relevant results, whether the authors synthesize across disciplinary boundaries or narrow the scope by focusing on specific journals, timespans, seminal works, or authors (Hiebl, 2023). Furthermore, scope considerations also need to include the reviews' envisioned method of literature analysis, whether it is quantitative (meta-analysis, bibliometric analysis, network analysis, etc.), qualitative (qualitative systematic review, integrative review, etc.), or a combination of both. Quantitative review methods use more isolated information from documents (e.g., effect sizes, citation links) compared to qualitative methods, which tend to be interested in larger portions of studies (Higgins et al., 2019). Arguably, authors of quantitative reviews will not need to pay so much attention to individual papers, and thus will be able to synthesize more documents compared to qualitative review authors.

The formal outcome of scoping is a review protocol describing concepts, the research problem and its significance, and inclusion/exclusion criteria (Tranfield et al., 2003). In medicine, this protocol is often preregistered at, e.g., PROSPERO (Daniels, 2019; Lakens et al., 2016). The so-called "scoping review" is considered a distinct review type in itself; it seeks to answer broader review questions and often forms a precursor to SLRs (Armstrong et al., 2011; Munn et al., 2018; Pham et al., 2014; Tricco et al., 2016). In management, the contents of the protocol are less formalized, as authors mainly set *"procedural boundary conditions [that] are concerned with curtailing the review's scope to specific periods, fields/subfields, journals, data sources, or databases"* (Simsek et al., 2021, p. 6).

4.1.1. Scoping means defining the search area based on a well-understood review topic

Accordingly, as part of the scoping phase, authors should build a strong foundation by tapping all available sources—including domain experts and the general internet—to learn about concepts and related language, define their topic of interest, and identify the first relevant studies. Domain experts, with their deep knowledge, can particularly expedite authors' familiarization with the research topic and point to important documents. To build a strong foundation for the subsequent searches, authors are advised to review contents and reference lists identified in seminal works and in previous and related reviews on the topic.

Optionally, issues of particularly relevant journals may be screened by hand (so-called "handsearching"), which is particularly effective when performed as early as the scoping phase. Identified relevant documents can then be used as inputs for the later systematic search steps. It is important to note that handsearching is not a search as such, but rather an exhaustive screening at a location where authors suspect many relevant documents. Systematic search heuristics (keyword searching, citation searching) aim to reduce the workload of such exhaustive heuristics while balancing the level of recall and precision.

Learning about a topic and discovery of first documents may involve any source of expertise, whether it be human intelligence or artificial, such as emerging large language model (LLM)-based discovery that may level the *"imbalance between the treasure trove of scholarly information and our limited ability to reach into it"* (Hope et al., 2022, p. 1). New advances in LLM-based search technology, such as ChatGPT, Scite.ai, SciSpace, Scopus AI, or Elicit.org, can facilitate the scoping process (Hutson, 2022), but these tools remain black boxes regarding search quality, transparency, and reproducibility. Accordingly, while such platforms are recommended for scoping, they cannot be recommended as principal tools for the later systematic search steps (Parisi and Sutton, 2024). Deep engagement in narrative searches from various sources facilitates authors' understanding of the topic and serendipitous discovery of relevant studies.

4.1.2. Frameworks to help define review scope

Inexperienced review authors may make the mistake of focusing their review efforts on overly broad topics such as *innovation* or *trust*. In such cases, we recommend narrowing the review scope by focusing on specific sub-topics or combining several topics. We can learn from evidence-based management that supports management decisions with the best available evidence. For this purpose, review questions must capture why or how the relationship occurs and in what circumstances (Denyer and Tranfield, 2009). The PICOC framework, an extension of the medical PICO framework, helps to formulate context-specific review questions (Barends and Rousseau, 2018). The acronym PICOC stands for Population (who is affected?), Intervention (management technique, method, independent variable), Comparison (alternative intervention, factor/variable), Outcome (objective, purpose, goal, dependent variable), and Context (type of organization, sector, country, other contextual factors).

Barends and Rousseau (2018, p. 25) give an example of a concrete review question for evidence-based management using PICOC: *"Do work-from-home arrangements (Intervention) lead to higher work satisfaction (Outcome) than office work (Comparison) for knowledge workers (Population) in tech companies (Context)?"* Including more PICOC elements would narrow the search scope, while excluding some would broaden it. Researchers might start with a few elements and then add more depending on the breadth of the review question and the number of identified studies. Crucially, omitting one or more elements from PICOC would yield a different review scope and answer; for example, work-from-home arrangements function differently for blue- versus white-collar workers (Population). PICOC's tight specificity means it will *"help to determine whether evidence will be generalizable and applicable to [an] organizational context"* (Barends and Rousseau, 2018, p. 25). Thus, the PICOC framework

includes considerations of the unit of analysis (individual, team, organizational, macro) that authors need to define.

While alternatives to the PICOC framework exist, such as the CIMO, SPIDER, ProPhet, or SPICE frameworks (Booth et al., 2016), not all SLRs will focus on narrow cause-and-effect relations. Authors may seek to synthesize broader topics instead and ask more general review questions, such as: *"What is the state of knowledge on work-from-home arrangements since COVID-19?"* Given the extraordinary breadth of such a review question, the resulting search strategy will be correspondingly broad, potentially encompassing an (overly?) large set of documents to screen and synthesize. Comparing these two very different example review questions illustrates the kind of decisions authors must make before systematic searching begins. Reviews with a clear purpose and narrow scopes that address specific issues with the greatest care will likely make it easier to sample diligently and conduct an impactful review (Simsek et al., 2024).

4.1.3. Limiting scope beyond concepts

While the review scope will be mainly determined by the (combination of) concepts of interest (e.g., elements of PICOC), the scope may be further limited using study characteristics. While medicine is relatively cautious about limiters such as document type, discipline, language, or geography (Konno and Pullin, 2020; Paez, 2017; Walpole, 2019), management research is traditionally more open. Overall, authors are free to determine the scope of the review by adjusting its research question, associated eligibility criteria, and database filters (Hiebl, 2023)—as long as such adjustments can be meaningfully justified.

Timeframe. Limiting time periods is a frequently used tactic, yet it can be challenging to justify unless the review topic provides a clear reason (Hiebl, 2023). Reasons that authors typically state include research developments, practice advancements, topic dynamism, time elapsed since a previous review, or some seminal work as the starting point. However, limiting the review periods to after a specific year may not lead to significantly fewer results, as publication output is growing exponentially, and most studies are published recently. For example, the Scopus database lists fewer management articles in the 104 years between 1900 and 2004 than in the six years between 2018 and 2022 (own analysis).

Document type. Limiting review scope to specific document types (e.g., journal articles, conference articles, theses, working papers, gray literature) is likely to narrow the number of eligible documents substantially. Limiting to peer-reviewed journals may increase the quality of the included documents, yet may also introduce publication bias (Dalton et al., 2012; Harrison et al., 2017), make the sample non-representative, or exclude emerging discussions. Inclusion of gray literature is mandatory for reviews that seek representativeness or that seek to increase statistical power in meta-analyses. In SLRs, gray literature inclusion is needed for *"new fields of inquiry where knowledge from early experience is needed"* (Adams et al., 2017, p. 447) and it is advisable when reviewers aim to include a *"richer, more detailed and practical understanding"* of the knowledge base (Adams et al., 2017, p. 446). To make the review as current as possible, authors should include unpublished studies, i.e., preprints. Databases' document type filters may be effective for narrowing the scope of SLRs (see Supplementary Material 1 or Search Smart).

Discipline. Subject filters limit documents to specific academic disciplines and may be helpful when dealing with ambiguous concepts that generate too many hits from disciplines outside the review's scope.

Journal. Limiting the scope to specific journals can be a powerful scope limiter, as it may include and exclude specific disciplines and journal qualities. Authors should be aware, however, that journal filters may limit representativeness or multidisciplinarity. While some suggest using journal quality thresholds for selection (e.g., Kraus et al., 2022), medicine and evidence-based management strongly advocate assessing study quality on a case-by-case basis (see Fig. 1: Phase 5). However, unlike medicine, where study designs such as randomized controlled trials are uniformly verifiable, management literature is more diverse and challenging to rate (Tranfield et al., 2003). The limitation of review scope to certain—often A-level—journals must be well justified, or it may have substantial negative consequences regarding diversity of ideas and sample size. By disregarding non-A journals, authors may solidify some of the negative consequences of the "an A is an A" mentality (Aguinis et al., 2020; Hiebl, 2023), leading to limited representativeness and unbalanced information (Cronin and George, 2023).

Citation count. Limiting reviews to documents with more citations—a sign of greater "influence"—is a practice occasionally used in reviews to cope with large samples (Hiebl, 2023). However, while citation numbers seem to be generally related to research quality (Thelwall et al., 2023), this relation does not necessarily hold at the level of individual studies. Thus, citation count filters will inevitably miss out on otherwise relevant, yet potentially unusual and innovative results (Thelwall, 2017) that SLR readers also expect to be synthesized. Both journal filters (often limited to A-level journals) and citation count filters, if applied without meaningful scope justifications, will contribute to a *"narrower scope of existing knowledge […] informing contemporary discovery and invention"* (Park et al., 2023, p. 142)—a trend particularly noticeable in the social sciences (Park et al., 2023). Selective synthesis of popular studies may contribute to a *"decline in disruptive science and technology"* (Park et al., 2023, p. 138). Thus, citation count filters need to be well justified (Hiebl, 2021) and, if adopted, applied in a methodically sound manner. This includes: (1) accounting for older studies being cited more often, and (2) accounting for discipline-specific differences in citation patterns ("field-normalization") if studies are sampled from multiple disciplines (for a simple explanation of both normalization approaches, see Waltman and Noyons, 2018).

Research design. Reviews may focus on specific research designs—like our own, which focuses on SLRs and MAs. Particularly, MAs will only be interested in empirical studies—primarily quantitative but also qualitative.

Language. The majority of documents in large databases are already written in English, but language filters can still be applied when necessary.

Geography. Although SLRs rarely use them currently, some databases provide location-based filters that can be helpful when it makes sense to limit the topic to specific locations. However, authors should beware of confusing publication location with study location.

4.1.4. Alternative scope-limiting approaches

Hiebl (2023) identified two alternative scope-limiting approaches: *seminal-work-driven* and *journal-driven*. The *seminal-work-driven* approach (Hiebl, 2023) is a unique and very narrow sampling approach where a highly relevant work located at the center of the review's attention, acting as a seed document, is connected to other works through citations. Review authors use citation connections (STEP 3) rather than language (STEP 2) to identify relevant documents. However, if the seminal work's influence extends beyond direct citations, a keyword search using its key terms may be necessary.

The *journal-driven approach* selects specific journals based on their suspected extensive coverage of a narrow review topic. This approach is particularly warranted when the review concerns a specific question pertinent to a specific journal or handful of journals—for example, how a certain discourse unfolded in a single journal.

4.1.5. Review scope determines the review burden

Once authors have decided on a specific SLR scope, sampling work will be judged based on SLRs' main quality criteria: comprehensiveness, transparency, and reproducibility. If authors cannot meaningfully limit the scope of their review, it is not acceptable to limit it with an incomplete search process; authors should choose a non-systematic reviewing method instead (e.g., Denyer and Tranfield, 2009; Paré et al., 2015; Simsek et al., 2021; Tranfield et al., 2003). To prevent the

unpleasant surprise of an overwhelming review burden, authors should familiarize themselves with search yields by conducting test searches as part of scoping.

In general, the review scope is at the discretion of the authors, who must meaningfully justify and report it. Scope-limiting choices must be transparently reported so readers can assess the representativeness of the review. For example, it is bad practice to claim that you have systematically reviewed a topic in the title and abstract, only to admit in a brief methodical remark that only a handful of journals were searched, or the period extends to just a few years.

4.1.6. Scoping outcomes

The outcomes of the scoping process are a well-defined review topic, a suitably narrow/broad search strategy, and a set of relevant documents identified through narrative searches, handsearches, or expert contacts. Similar to the documents that are identified in the later steps, the initially identified documents are most economically screened in two steps—first title/abstract and then full text (see Table 1). Even in the scoping step, authors already need to familiarize themselves with the reporting requirements of searching and screening (see STEP 4) so they remember to note important information that is already relevant.

Table 1
Step 1 - scoping the review (mandatory).

Procedural details	Tools/help	Evaluation questions
Learn about the topic of interest • **Only if the topic is well understood** should you embark on the actual systematic search (STEPS 2–3). • **Learn about the topic** via: (1) narrative/exploratory searches of the internet. Actively seek serendipitous discovery of concepts, moderators, and language variations (2) reading important works and previous or related literature reviews on the topic (3) asking experts. • **Build on others' work.** You may find entire search strings useful to your search strategy. Searching within related reviews (MEMR5-mandatory) for relevant documents, concepts, or keywords is mandatory.	• Use popular sources, such as *Wikipedia*, or narrative *Google (Scholar)* searches. In particular, semantic search engines like *GS* or *Semantic Scholar,* or new *LLM search systems (*e.g., *SciSpace, Elicit, Perplexity, or Scite)* might help expand your searches beyond the language you are familiar with (Hutson, 2022). Be cautious with AI-generated summaries, as many may be entirely fabricated. • New research ranking metrics might help uncover particularly popular articles within a topic. Use *Dimensions.ai* to search for topics or within specific journals and rank via Altmetrics (research attention) or Citations (research interest).	• Is the review topic clearly defined? • Does the SLR address a clear review question? • Are prior reviews acknowledged and used to inform the SLR's scope and identify first relevant studies?
Identify first relevant documents • Continually add references to the **initial set of narratively identified relevant documents.** Report documents you find narratively in the search protocol (STEP 4). • **Contact authors** well-versed in your review topic and **ask for unpublished studies** (MEMR7-highly desirable). • Use optional **mass-solicitation or open calls** to seek works.	• For specific terms that mainly appear in the full text and not in the title, abstract, or keywords, a *GS* search of the full text using a hardcoded search string (terms enclosed in quotation marks) might be helpful. It has the best full-text coverage, yet limited Boolean capabilities for accessing it (for STEP 2).	• How many relevant documents are identified during scoping?
Build your initial search string by learning about language • **Build your initial search string** by continuously analyzing and **noting important concepts and keywords** (which concepts are important to delineate what is relevant from what is irrelevant for your analysis? What different terms/keywords are used to describe each of those concepts?) • **A well-designed, comprehensive search string is mandatory.** Account for language variations used by different authors describing the topic. Include all important concepts, synonyms, related terms, spelling variations, and acronyms, combined with meaningful Boolean operators (see STEP 2 for details).	• Extract keywords from identified documents or bibliographic text files via multidisciplinary text-mining tools such as *Text Analyzer* and *Voyant* (O'Keefe et al., 2023). • Expand your already-identified keywords with *online thesauri* or the *controlled vocabularies* supported by some databases (e.g., Business Source Premier or ABI/Inform Global) or *LLMs* to suggest terms.	See STEP 2
Define eligibility criteria • Continually (**re-)define eligibility criteria** and make them explicit (what criteria must a study fulfill to be relevant for your topic of interest?). The search protocol, the review's scope, its research question, and eligibility criteria must be clear before STEP 2. • Eligibility criteria also determine the selection of suitable filters to meaningfully limit the searches.	• For evidence-based management topics, use frameworks such as *PICOC* to scope your review topic meaningfully and to define relevant concepts (Barends and Rousseau, 2018).	• Are inclusion/ exclusion criteria defined explicitly and well justified?
Test the keyword search strategy (search string and filters) • **Perform first keyword searches** to test whether the number and ratio of relevant/irrelevant studies identified correspond to your envisioned review scope and synthesis depth. You might want to calibrate (1) the search area via more or less broad search strings and (2) the scope of your research question/topic. • Databases envisaged for STEP 2 should be already tested in STEP 1. Find out about the size of the literature base of your scoped review topic.	• Ask a specialized *librarian* or similar information specialist to help with designing search strings. Research has found that their help (and co-authorship) noticeably improves the quality of SLRs (Meert et al., 2016; Schellinger et al., 2021). • *LLMs* may help to design the first search string and detect errors, yet LLM-based strings will likely remain mediocre and incomplete (Wang et al., 2023) and rarely use truncation, wildcards, or proximity operators. Human expertise is still needed (Parisi and Sutton, 2024).	See STEP 2

(*continued on next page*)

Table 1 (*continued*)

Procedural details	Tools/help	Evaluation questions
Handsearching • **Handsearching** is the manual screening of all documents in a specific location that authors have identified as relevant—most often in a specific journal (issue) but can be specific authors or institutions that focus on a topic. Handsearching is a search heuristic that does not perform searches but requires results to be screened for relevance: for example, every document within a specific journal, a database, or a document's citation data (Cooper et al., 2020b). It should be performed if a highly relevant outlet is identified (many documents are suspected to be relevant there). • Handsearching is rarely the sole search heuristic used for a review (unless the review scope is so narrow that it only includes the journal/author/institution being handsearched). Handsearching would usually supplement keyword-based methods, particularly if a specific journal is highly relevant for the review and many hits can be anticipated. • If you choose to handsearch to supplement other heuristics, it is most effective to do so before all other search steps. Handsearching results can then inform scoping decisions and the subsequent search steps: keyword, forward, and backward citation searches will benefit from the documents/ keywords identified initially.	• Handsearching can be done by browsing issues on publisher websites (e.g., *SpringerLink, Wiley*; many are listed in Supplementary Material 1) or by using journal field codes at bibliographic databases that support them. • You can facilitate and organize handsearching of multiple journals via a database search of multiple journal identifiers (ISSNs). It is important to verify in advance that the database covers the journals in question.	Handsearching is most often optional
Screening of identified studies (perform throughout STEPS 1–3) **Decide whether the identified studies meet your inclusion criteria during screening**. Screening is most economical if done in steps: **1) Deduplication** of documents across databases and search methods to save screening time. The documents should be saved using reference management software to track which ones were retrieved at which stage and to organize them throughout the process. Saving all articles without deletion ensures effective deduplication in later steps. **2) Screening 1 (based on title/abstract).** Screening documents first on title/abstract information saves the time needed to download irrelevant full texts. In cases where relevant information is not in the title or abstract, more full texts will need to be screened (see "difficult-to-search review topics"). Final decisions should be made based on full-text information. If in doubt, keep the document included and make your final decision with the help of the full text in the next screening step. **3) Screening 2 (based on full text).** Full texts of the studies that were not deemed irrelevant in the first screening round (positives and question marks) are retrieved. Final inclusion decisions should be made based on full-text information. If full texts are unavailable or key information/data is missing or unclear, study authors must be contacted before studies are excluded. Particularly for meta-analysis, it is important to consider whether such exclusions may introduce bias to the analysis (MEMR15: mandatory). Screening decisions (including exclusion reasons) must be documented, ideally in a flow diagram (see STEP 4). A table of otherwise relevant studies that were excluded solely due to missing key information should be documented (MEMR16: mandatory). In this step, studies with sample/content overlap should be merged, or one of them excluded, to represent each only once (Wood, 2008) (particularly relevant in MAs) (MEMR17: mandatory). **The necessity of two screeners:** Two screeners should independently perform screenings of titles/abstracts (STEP 1) and full texts (STEP 2) (MEMR14: highly desirable). However, in the case of extraordinary large search result sets that show extraordinary diligence in covering the search area, one reviewer may perform the initial title/abstract screening and two or more screeners perform the full-text screening. Careful and in-detail specification of multiple objective exclusion criteria have tend to improve screening accuracy to similar levels as multi-person screeners (Nama et al., 2021). For 2+ screeners, inter-rater reliabilities may be calculated (Hiebl, 2023).	• Use *reference management software* to save references and track which documents were retrieved at which search (important for reporting – see STEP 4) • *Zotero* or *EndNote* may be used to automatically download available full texts. • *Systematic Review Accelerator's Deduplicator* was effective in deduplication and can be used as a standalone or add-in in Endnote (Guimarães et al., 2022). • Popular tools to support the screening process through decision management, screening organization, and data handling are *Rayyan, DistillerSR, Abstrackr, or Covidence* (Khalil et al., 2024). • *LLM-enabled tools such as ChatGPT* may be used as an additional screener or as an *"efficient first-line screening tool"* to exclude more obviously irrelevant documents for well-structured topics, such as in medical papers with structured scopes (Issaiy et al., 2024, p. 78). Reviews with more complex concepts are less likely to benefit from AI-assisted screening.	• Are the inclusion/ exclusion criteria used to rate study eligibility? • Is the screening process transparently reported (incl. the exclusion decisions at each screening step)? • Do two authors independently screen the studies and in case of a single screener was such an approach meaningfully justified?

(*continued on next page*)

Table 1 (*continued*)

Procedural details	Tools/help	Evaluation questions
Rating approaches for 2+ screeners: One approach to rating the suitability of articles more granular than "include/exclude" is a points-based logic, where inclusion is determined based on the mean score assigned by the screening authors (Hiebl, 2023). Additionally, the A/B/C logic rates articles as A (particularly relevant items), B (potentially relevant items), or C (items with little or no relevance) (Pittaway et al., 2004).		

Box 1
Comparing STEP 1 to current scoping practice in management SLRs and MAs.

Management SLRs and MAs in our sample were authored by 3.65 researchers on average (median 3), with a minimum team size of one and a maximum size of 16. Most reviews base their search strategy on a database-driven logic (96%), supplemented by other heuristics to ensure comprehensiveness. However, only a few studies reported dedicated narrative searches (14%) preceding keyword searches. Additional steps in the scoping stage, such as handsearching individual journals and relevant reviews (23%), along with expert consultations (19%), are also reported infrequently. These low percentages highlight reporting deficiencies at this stage and also significant shortcomings in the scoping procedures that precede systematic search phases (STEPS 2–3). Effective scoping through thorough identification and screening of relevant language and documents is critical for successful review outcomes and should also be reported. Only a well-scoped topic ensures that subsequent language-based (keyword) and associative (citation) searches are representative of the topic.

Review authors usually apply database filters to limit the scope of their searches to fit inclusion or exclusion criteria. We found that 349 studies (86%) applied some filters along with the keyword searches they performed. More specifically, we found that 156 articles (39%) limited the journal quality; for instance, several reviews only included articles published in peer-reviewed journals in their sample. A total of 104 studies (26%) limited the study type (e.g., journal articles, book chapters), 224 studies (55%) limited the language, and 136 (34%) limited their searches' timeframe. Ultimately, we found that only one article limited the search to a specific geographical scope. The common reliance on database filters to refine the search shows the authors' commitment to search precision. To do so, authors need to know at which databases such filters are supported (see Supplementary Material 1).

Authors should only move on to searching systematically for relevant documents once they have a clear comprehension of the review question (s), core concepts, and search breadth.

4.2. STEP 2: Systematic keyword searching (for details, see Table 2)

After thorough scoping, review authors should feel confident searching systematically for relevant documents in suitable bibliographic databases. Designing a comprehensive search strategy, including search strings and filters tailored to the databases to be searched, is mandatory.

4.2.1. Search string formulation

Keyword searching identifies potentially relevant documents via associations of language expressed through search strings (rather than associations via citations in STEP 3). Through scoping, authors will have defined a meaningful review question and identified essential topics and the language in which they are described. However, authors from different backgrounds often describe the phenomena of interest in different language. Thus, review authors need to go beyond the terms they are familiar with and collect concepts' different meanings by, for example, extracting keywords and synonyms from identified relevant documents.

Reviewing definitions of topics helps to clarify these concepts and identify different terminology. For example, in screening documents on the concept of *offshoring,* some authors define it as: *"the relocation of business functions from home base to foreign locations"* (Roza et al., 2011, p. 314). This definition shows that the concept consists of two (or more) distinct sub-concepts: "relocation" and "from home base to foreign location." Studies may have used different language to convey the same meaning for each sub-concept. They might call it "transfer" and/or "abroad"—two of the many synonyms review authors should include in their keyword search. Already-identified keywords may be expanded via *online thesauri,* the *controlled vocabularies* supported by some databases (e.g., Business Source Premier or ABI/Inform Global), or *LLMs.*

Search strings in keyword searches are primarily formulated using the "building block" logic (Booth, 2008), where single concepts are combined with Boolean operators to ensure they jointly span a search area. Each concept staked out in the scoping phase (e.g., a PICOC element) will become one block of synonyms combined with Boolean "OR" operators and truncation and wildcard symbols to include word variations. These blocks are then combined with Boolean "AND" operators so they jointly delineate studies on the topic from those outside the scope. The search area should be as wide as necessary to encompass the most potentially relevant documents while the share of irrelevant documents remains acceptable. While the "OR" operator should be used liberally, review authors must be wary of the "AND" operator, as including too many concepts at once would overly limit the search scope. Keyword searching must always consider the tradeoff between precision (the least irrelevant documents) and sensitivity (the most relevant documents). Precision and sensitivity must always be balanced: perfect precision has the disadvantage of missing out on relevant results, while perfect recall has the disadvantage of escalating workload.

4.2.2. Two quality checks: Validation of the keyword search strategy and search breadth

We suggest two quality checks to discover red flags in the keyword search strategy. These checks should be applied after thoroughly conducted scoping and finalizing search string composition. Check #1 helps identify non-representative search strategies, while check #2 helps identify excessively narrow and broad search strategies; all these issues pose limitations to sampling effectiveness. Only if both checks are

passed should authors proceed to systematically retrieve results through keyword searches. The results of these checks should also be reported (see STEP 4). It is important to remember that the checks are necessary, yet not sufficient signs of search diligence—meaning that some other deficiencies may still remain undetected.

Quality check #1: "Validation of the keyword search strategy." This check validates whether the keywords and their composition in the best/last search string are capable of identifying all the documents identified during the scoping phase (e.g., Pullin et al., 2018). The informative value of this test depends on the thoroughness of the initial scoping (narrative searches and handsearches). A greater number and diversity of narratively identified relevant documents allow for a more meaningful assessment of search string design. The procedures for check #1 are detailed in Table 2 and an example is illustrated in Supplementary Material 5.

Quality check #2: "Validation of keyword search breadth." The "number needed to read" (NNR) indicator helps evaluate review diligence by identifying excessively narrow or broad sampling scopes (Cooper et al., 2018b). The NNR is the number of documents a researcher must read (screen) to identify a relevant one. While methodologists are careful not to determine a specific NNR threshold for SLRs (Cooper et al., 2020a), well-tailored search strategies with meaningful database filters can produce good results with NNRs as low as 8.3 or 20 (Bachmann et al., 2002). Bramer et al. (2017) reviewed different database combinations in medical SLRs and found that sensitive searches (overall recall above 80 %) had NNRs between 43 and 73, with a NNR of 73 identifying the most relevant documents.

Similar to medicine, we do not advocate a single threshold, but rather a careful consideration of all elements in the sampling process. Based on what we know from medicine, a meaningful rule of thumb can flag (overly) low NNRs (under eight) that will likely reflect low search sensitivity (many relevant studies probably omitted), while overly high NNRs (above 100) may imply imprecise searches due to searching in the wrong places or with the wrong language or techniques. Thus, while a NNR value outside this band likely indicates issues with search quality (as we will show later), a value within the band does not in itself constitute a reliable indicator of good search quality. It is important to note that NNR values can only be assessed jointly with the authors' review scope and overall search strategy to give a good picture of sampling quality (Ross-White and Godfrey, 2017). Based on the NNR review, authors can more accurately assess search quality and peer reviewers and editors may be able to give more targeted suggestions for sampling improvements. The procedures for check #2 are detailed in Table 2.

4.2.3. Choice of principal databases: Considering keyword search functionality

Searching with a few highly suitable databases is likely just as important for the quality of keyword searches as developing high-quality search strings (cf. Booth, 2010). Suitability is determined by both document coverage and search functionalities. While it is essential that databases cover many relevant documents, this does not necessarily mean that those documents are also *discoverable*. Many search interfaces limit or impede the discovery of these otherwise covered documents (Gusenbauer and Haddaway, 2020). Only a fraction of databases support keyword citations without severe limitations—even fewer than support extensive download options facilitating the sampling process. One database that performs particularly poorly at keyword searches is Google Scholar (GS), where misinterpretations of Boolean search queries lead to relevant results being omitted. Consequently, if researchers rely on databases with such inadequate capabilities, their search results may be either non-retrievable, non-reproducible, inaccurate, or lacking transparency—all serious issues in systematic searching, the goal of which is comprehensiveness, transparency, and reproducibility.

To illustrate these issues, we replicated Maddux et al.'s (2021) reported GS search strategy. We were puzzled that while GS covers all the documents their study analyzes, just 15 % were identifiable via the search strategies they reported.[2] At least two functionality deficits limit this discoverability: GS's faulty interpretation of Boolean strings leads to partial search coverage of relevant concepts, while its display limit of 1000 search results prevents authors from accessing (not to mention downloading) all potentially retrieved results. Thus, if a database's keyword search functionality is *"limited,"* it should not be used as the *principal* database.

Supplementary Material 2 illustrates additional noteworthy capability differences. While some databases allow queries with >9000 keywords (e.g., Web of Science), others substantially limit search string length: Ingenta Connect, ScienceDirect, Semantic Scholar, and SpringerLink allow only up to 10 keywords; JSTOR, Open Access Theses and Dissertations, and SSRN allow about 25 keywords; ERIC, DOAJ, and EconPapers allow about 50 keywords; and GS has a strict limit of 256 characters. These limitations are one reason why systematic keyword searches can be challenging or, in some cases, impossible with certain search systems. Knowing such limitations in advance is important for review authors to target suitable databases from the outset and tailor their search strategy accordingly.

4.2.4. Choice of principal databases: considering coverage

In addition to database functionality, which often presents hard knock-out criteria for selection, the extent to which databases cover relevant management literature determines search prioritization (Wanyama et al., 2022). In line with the Cochrane Handbook, we distinguish between larger multidisciplinary (higher *number* of management documents) and smaller specialist databases (higher *share* of management documents). Lens, for example, is a large multidisciplinary database covering more than six million management documents (high *absolute* coverage) along with many documents from other disciplines (>97 % are not management-related). Emerald Insights, however, is a specialized database that primarily covers management documents (with high *relative* coverage at 49 %), yet comparatively fewer (with under 300,000 management documents, meaning low *absolute* coverage).

Large multidisciplinary databases are the principal information sources for every systematic search strategy (Higgins et al., 2023) and are even more vital when review questions target multiple disciplines and are written in the form of an integrative review to synthesize across research approaches (Cronin and George, 2023). One of management's primary review journals, the *International Journal of Management Reviews*, explicitly emphasizes this interdisciplinarity and notes that coverage issues are the second most common rejection criterion (O. Jones and Gatrell, 2014). Accordingly, choosing large, multidisciplinary databases with high absolute coverage of management literature is essential.

4.2.5. Choice of principal databases: considering search filters

Further, review authors should employ filters reflecting their previously defined inclusion/exclusion criteria during keyword searching to increase search effectiveness. Databases support various filters that allow authors to narrow search results before or after a keyword query is performed. Authors may filter for e.g., 1) publication years; 2) publication outlets (i.e., via the title or a unique identifier such as ISSNs); 3) subjects (management vs. other disciplines); 4) document types (e.g., journal articles); 5) languages; or 6) geographies. However, not all databases support all filters (for an overview of which filters are supported,

[2] We used the Publish or Perish download function on 95 of the 145 references reported by the authors.

consult Supplementary Material 1 or *Search Smart*). The advantage of employing meaningful filters is improved search precision (a higher share of relevant documents with the results set) and, thus, a reduced number of documents that must be screened for every relevant document. If the review scope permits restrictive filters, using them appropriately could reduce the screening workload by 80 % or 88 % (Ayiku et al., 2020). However, inappropriate filters will omit relevant studies and reduce the scope of the review (Leeflang et al., 2006).

4.2.6. Selecting suitable principal databases with Search Smart

Review authors need to use databases whose systems adequately support search functionalities for keyword searches (STEP 2) or citation searches (STEP 3), while covering most of the desired documents (for an overview, see Supplementary Material 1). To identify particularly suitable databases, we use Search Smart (Gusenbauer, 2024b), a website that evaluates and ranks all major management-related bibliographic databases. While we describe the database selection process here, readers can directly replicate and update the steps at the website as well.

At Search Smart, because we only want to consider databases supporting keyword searches with Boolean logic and that support bulk downloads, we select the *"Systematic keyword searching"* preset with the *"Bulk select & export options"* setting. To rank databases with highest management coverage at the top, we select *"Subject coverage > Business, Management and Accounting"* and the *"Subject coverage (abs. / desc.)"* sorting. The resulting list shows that ABI/Inform Global (via ProQuest), followed by Lens and Business Source Premier (via EBSCOhost), are the most comprehensive management databases. If authors seek journal articles, it is important to bear in mind that not all databases cover them equally (e.g., ABI/Inform Global and Business Source Premier/Elite, etc. only cover 13.38 % and 16.70 % respectively). Thus, for searching journal articles, Lens will be the better option, followed by Dimensions. Another consideration is the proportion of other relevant document types, such as conference proceedings, theses, working papers, and gray literature (Adams et al., 2017). Particularly, if authors focus on higher quality peer-reviewed articles only, they will be interested in WOSCC (full coverage) with a slight advantage over Scopus as one of their principal databases—despite their disadvantages in terms of absolute coverage.

Additionally, authors should use specialized databases with relatively extensive coverage of the disciplines their review touches upon (e.g., management, computer science, psychology) or populated with certain document types (e.g., theses, conference proceedings, gray literature). At such specialized databases, searches will be much more precise, as they contain fewer irrelevant—i.e., non-management related—results. At the same time, recall will be inferior to that of large, specialized databases. The levels of precision and recall of searches will depend on how well the database fits the review topic. For example, if authors' review topic touches on psychology, authors may use Search Smart to select *"Subject coverage > Psychology"* and the *"Subject coverage (rel. / desc.)" sorting*. The resulting list shows APA PsycNet, followed by APA PsycInfo (via EBSCOhost and Ovid) and Psychology & Behavioral Sciences Collection (via EBSCOhost) as the databases most specialized in psychology.

To illustrate the tradeoff between large, multidisciplinary, and specialized databases, we analyzed the document coverage of two exemplary psychology-related SLRs (n = 108; Maddux et al., 2021; Yeung et al., 2021) at the 50 management-related databases. We found that the large, multidisciplinary databases covered most of the articles, and GS and Lens even covered all of them. Smaller specialized databases covered fewer, and the publishers' databases covered only a fraction of the documents. The example of APA PsycInfo, a small database specialized in psychology, shows that specialized databases are only valuable when they match the review topic. Despite its significantly smaller size, the database identified the largest share of documents, since the databases and review topics' foci matched. The key takeaway is that larger, multidisciplinary databases should be the foundation of keyword searching, while smaller, specialized databases support the larger ones in the relevant disciplines they cover.

4.2.7. Number of principal databases to search with

In a perfect world, searching one database containing all relevant documents (full coverage) with a perfect search query (full recall with perfect precision) would be sufficient. However, the reality is that no single database covers all relevant documents, and the one that comes closest (GS) (Gusenbauer, 2019) does not support Boolean queries and unfortunately has other limitations too (Gusenbauer and Haddaway, 2020). Therefore, it is recommended that researchers search multiple databases with different search heuristics (keyword/citation searches) to counter databases' limited coverage and retrievability and the limited human capacity to articulate well-designed search queries (C Lefebvre et al., 2023; Salvador-Oliván et al., 2019).

While there is likely no definitive optimal number of databases to search (Stevinson and Lawlor, 2004), there seems to be a consensus that a mix of a few highly relevant multidisciplinary and specialized *principal* databases is sufficient for an effective keyword search (Aagaard et al., 2016; Bramer et al., 2017; C Lefebvre et al., 2023). **As a rule of thumb, we recommend using *at least* three highly suitable databases for keyword searches—at least two large, multidisciplinary ones, and at least one specialized database for each larger subject the review topic touches on.** Three databases (as a minimum) for keyword searching echoes the suggestions of Beyer and Wright (2011), Frandsen et al. (2019), and Rousseau (2024) and reflects the median number of searched databases from our and Hiebl's (2023) analyses. Keyword searching two large multidisciplinary databases leads to higher recall and may also hedge against authors making database-specific query mistakes (Salvador-Oliván et al., 2019). Further, one or more specialized databases for each subject are meaningful, as specialized ones tend to identify unique documents (Bramer et al., 2017). Further, authors should add one or more databases specialized in specific relevant document types (e.g., gray literature, theses) in case they seek documents barely covered in the principal databases. Additional citation-based searches (STEP 3) will facilitate the identification of difficult-to-search topics, which are particularly prevalent in management research.

To identify particularly suitable databases for other disciplines, authors should consult the *Search Smart* website (Gusenbauer, 2024b).

4.2.8. Supplementary databases

In addition to the *principal* databases the review should rely upon, Cochrane Handbook guidance (Higgins et al., 2023) encourages the use of *supplementary* sources that may even be functionally unsuitable. The logic is that while searching more databases will never hurt, it might identify *"'unique' studies that are not identified by other search methods"* (Higgins et al., 2023, p. 25).

4.2.9. Updating the search process

From the beginning, review authors should consider that they might need to update sampling—particularly in management, with a peer review process that lasts twice as long as that of medicine and is the longest among all disciplines (Huisman and Smits, 2017). The Cochrane Handbook advises that the possible need to incorporate new results 12 months before publication should be weighed against the risk of publication delay.

Table 2
Step 2—systematic keyword searching (mandatory).

Procedural details	*Tools/help*	*Evaluation questions*
Finalize your last/best keyword search strategy (MEMR8/ MEMR9-mandatory) **Calibrate your search string** by assessing its preliminary search results (yield): • One or multiple concepts span the topic's search area. Each individual concept should contain all relevant terms, synonyms, and wordforms (using truncation and wildcards) to capture the ways authors (could have) described it. **Keywords within concepts are combined with the OR operator. Multiple concepts are combined with the AND operator.** Adding concepts will limit the topic's scope and search area. It is recommended to start with core concepts only (related terms connected with OR) and then add additional ones (with AND) if the results set is excessive and if the added concepts are important delineations of the review scope (Higgins et al., 2023). • Reflect on where in the document the information will appear and choose suitable field codes (traditionally title/ abstract/keyword field codes, but could also be the full text, if databases support it). • **If the search yield is too low:** - broaden the review/search scope by removing a concept or an eligibility criterion (i.e., database filter). - increase the number of relevant synonyms, etc. for each concept. • **If the search yield is too high:** - narrow the review/search scope by adding an additional concept. - validate the yield of single terms and potentially exclude single highly ambiguous terms. Occasionally, common and ambiguous terms are responsible for many irrelevant results. • **If searches are imprecise (high NNRs):** - try using **proximity operators** to identify terms that are proximate but not necessarily directly adjacent or in the same order. Compared to the AND operator, which does not consider word distance, a proximity operator—if used diligently—may considerably improve search precision. - try adding **subject limiters** via databases' subject filters (see Supplementary Material 1) or by adding another relevant subject-related concept that disambiguates the search string. ATTENTION: this will also limit cross-disciplinary hits—a scoping decision that needs to be reported in the review's scope. - **Generally, be careful with the NOT operator**, as it will likely also remove relevant documents. Test searches of excluded results help determine whether the NOT operator has been used appropriately. In other words, exclusively search for the excluded results and review them for relevance. • **Validate your search string with the quality checks #1 and #2 (see below) and the PRESS 2015 Evidence-Based Checklist.** PRESS is tailored to medicine, yet all items (except subject headings) are highly relevant for management too (McGowan et al., 2016). • After iteratively testing your keyword search—also across the databases you selected—you should have developed a comprehensive and well-calibrated search string. With this string, perform queries, retrieve results, and document the process (see STEP 4).	• Search string formulation is also explained simply in the *Evidence-based management book* (Barends and Rousseau, 2018, pp. 110—136). • For further sampling help *consult specialist librarians or other information specialists,* whose assistance has been shown to improve SLR quality (Meert et al., 2016; Schellinger et al., 2021).	• Is the search string representative of the diverse language of the review topic? • Does the search string include a representative selection of keywords? • Does the search string correctly apply Boolean operators and field codes to cover the entire search area? • The *Peer Review of Electronic Search Strategies (PRESS) 2015 Evidence-Based Checklist* provides a list of questions to evaluate keyword search strategy elements.
Specify filters • Along with your inclusion/exclusion criteria, determine useful filters that help you further narrow down your results set and increase search precision. Database filters must be well justified and align meaningfully with your review's research question (Hiebl, 2023) (MEMR10: mandatory). Popular filters include publication year, subject coverage, document type coverage, publication outlet, and language (for availability, see Supplementary Material 1).	• Use *Search Smart* (Gusenbauer, 2024b) to identify which databases support which document types and filters.	• Do database filters align with the predefined inclusion/ exclusion criteria?

(*continued on next page*)

Table 2 (*continued*)

Procedural details	*Tools/help*	*Evaluation questions*
• The choice of document types reviewed is at the review authors' discretion and must be determined in the review scope (Adams et al., 2017). Reviews should include a meaningful selection of document types. Specifically, if reviews aspire to representativeness, they should search beyond journal articles, including conference proceedings, working papers, theses, and gray literature (MEMR3: highly desirable). MAs will be interested in gray literature to increase statistical power. SLRs will include gray literature to increase the search yield of under-researched topics, investigate emerging discourses, improve practical relevancy, or to improve representativeness (MEMR4: highly desirable). • Journal quality, often used as a (questionable) proxy for study quality, may be limited by using ISSN journal identifiers to narrow keyword searches to a list of selected journals that meet specific inclusion criteria. Because of their unique IDs, ISSN filters are likely better at identifying journals than just using journal names. Supplementary Material 1 indicates where searches via ISSNs are currently feasible.		
Quality check #1: keyword search strategy validity Before you start with the systematic keyword searches, **validate/live-test your keyword search strategy against your already-identified set of relevant studies from STEP 1 (via narrative searches and handsearching of existing reviews and journals)**. The more relevant studies exist, the more meaningful this check will be. The goal is to test **whether your last/best search string identifies all documents at one of your principal databases** (the one with most absolute subject coverage). 1) **Compile your last/best keyword search strategy** (including search string, field codes, and filters the database supports). 2) **Compile the set of already identified relevant studies at the database**. Do this by searching for individual documents (via DOIs or document titles). Not all documents will be identifiable via DOIs (and not all databases support DOI searches), so you may need to perform title searches. With title searches, be careful with dashes (–), hyphens (-), and apostrophes (`), which are frequently misinterpreted by search systems (Zhou et al., 2021). Combine the individual DOIs and titles with OR operators to receive a results set comprising only relevant studies. CONSIDER: Some studies may not be covered at your principal database. In this case, you can perform the same steps at some other relevant database or omit these studies from the test. If many studies are not identifiable, you may be searching the wrong database (this will also depend in study type – e.g., gray literature will be harder to identify at major databases like Scopus or WOSCC). 3) **Compile a Boolean search string that combines: 2) NOT 1)** If your search string 1) identifies all documents from 2), then 3) will show zero results. The results still shown will be those your search strategy does not identify, i.e., that lie outside your search strategy's scope. 4) **Review why the remaining documents were not identified.** Which keywords are missing? Do Boolean operators function as intended? What field codes are overly limiting? Are filters too restrictive? etc. Adjust your search strategy accordingly so all relevant documents will be identified and weighing search recall against search precision. **Consider:** The check reveals how well the search string covers the language of documents already identified (during scoping), yet not how far the string's language extends beyond that. Thus, the check will only be meaningful if scoping was thorough and exploratory, yielding many relevant documents.	• An example of quality check #1 is illustrated in Supplementary Material 5.	• Is the keyword search strategy validated, such that it finds all studies identified in scoping?

(*continued on next page*)

Table 2 (*continued*)

Procedural details	*Tools/help*	*Evaluation questions*
Quality check #2: keyword search breadth validity Apply this check as you start searching systematically • 1) **To get a rough idea** of keyword search precision (inverse NNR), compare the number of studies your string identifies from your already-identified studies during scoping (quality check #1) to the total number of results it identifies. This is the upper bound of your NNR (if no additional relevant studies are found beyond those already identified). The NNR will likely decrease substantially as more articles are rated as relevant during screening. • 2) Within the results set you are about to screen in full (after you have already applied all database filters), screen a first random sample to estimate the NNR of this search (i.e., the estimate of how many documents you need to screen to identify one relevant one). • 3) If the number of irrelevant hits in your keyword search is too large (high NNR), you may need to make your search string more precise (see procedures above). If the number of irrelevant results is relatively low (low NNR), you might miss out on many relevant results due to your overly precise search string and need to broaden the search strategy. **NNRs below eight or above 100 should be subject to particular scrutiny; however, an NNR within these bounds does not constitute a *guarantee* of a good search strategy in itself**. • 4) After screening of keyword searches is completed (STEP 2), report the final NNR—i.e., divide the number of screened studies by the number of relevant studies. Only consider those studies that were identified through keyword searches, and not other search methods (see STEP 4).	• Further details on interpreting the NNR can be found in the gray box below.	• Is the NNR (screened studies divided by relevant studies) between eight and 100? • Is a NNR below eight meaningfully justified?
Select suitable *principal* databases **Select two or more principal databases with high *absolute* subject coverage** (MEMR1: mandatory)**:** • Select two or more large, multidisciplinary databases that support Boolean searching and cover the most documents in your subject(s) of interest. • High absolute coverage (e.g., *Lens/Dimensions* for various document types, or *WOSCC/Scopus* for focus on higher-quality types) is advantageous for sensitive searches. **Select one or more principal databases with a high *relative* subject or document coverage for each discipline the review touches on** (MEMR2: mandatory): • Select one or more specialized databases for 1) each main subject of interest your review touches on (e.g., management, psychology, computer science) or 2) each specific (non-journal article) document type the review includes (conference publications, unpublished studies, gray literature, etc.). • High subject specialization (e.g., *ABI/Inform Global* or *Business Source Premier*) is advantageous for low-precision queries that span multiple (irrelevant) subjects. Searching specialized databases is particularly important if additional document types must be considered (e.g., dissertations, gray literature, conference papers, and preprints). **Additional considerations for database selection:** • **Coverage of non-English documents:** most major literature databases have an English focus. You might want to/must also consider non-English databases for your review. • **Coverage of newly published documents:** consider databases that support "articles in press" or "early online access," or search for preprint versions at dedicated databases that publish such document types. • **Resource constraints (open discovery):** Not all databases will be available to you. Use non-paywalled tools offering equal performance in terms of coverage and functionality. While these open tools give free access to the search interface and citation info, review authors must still obtain access to a large volume of paywalled documents. *Google (Scholar)* searches will identify many free versions of the paywalled documents. Review authors should contact the authors of the remaining missing studies for a copy.	• Use *Search Smart* (Gusenbauer, 2024b) to identify suitable principal databases. • Use *Systematic Review Accelerator's Polyglot Search* to translate search strings between *Scopus, WOS, and PsycInfo (Ovid)*.	• Do authors search *at least* two highly relevant multidisciplinary databases and at least one specialized database for each covered main subject?

(*continued on next page*)

Table 2 (*continued*)

Procedural details	*Tools/help*	*Evaluation questions*
Select *supplementary* databases **Search supplementary databases, *in addition* to principal databases:** • **Some databases have severe limitations on Boolean searching.** Consequently, we advise against their use as *principal* databases for keyword searching. However, *additionally* searching more databases, even unsuitable ones, will never harm search yield. Searches of non-suitable academic databases (e.g., *GS*) or even popular search engines (e.g., *Google, Bing*) and LLM-based searches may be used to *supplement* (though never replace) the principal databases and might identify some unique studies not identified elsewhere. If appropriate, such searches should use the same search terms as the other keyword searches to ensure consistency (Higgins et al., 2023) and should also be reported similarly to the principal databases. • Weigh the additional workload, bias, and non-reproducibility these searches introduce against the additional documents that will potentially be identified. • In general, searching more databases will improve the result, yet with decreasing rates of return. **It is often better to improve the search string and perform both backward and forward citation searches than to add databases in the absence of one of the reasons above.** Signaling diligence by searching many databases with inferior strings and overly restrictive filters is not good practice (Booth, 2010).		• If supplementary databases are searched, are they reported well?
Updating the search • **During the search process:** If significant new terms (language) appear throughout the sampling process, STEP 2 needs to be repeated. • **After the search has been completed:** The SLR should be as up to date as possible. MAs, in particular, may benefit from larger samples due to updating. Whether the searches should be re-run within 12 months before publication needs to be evaluated on a case-by-case basis by editors and peer reviewers (MEMR12: highly desirable). Whether to add new documents must be weighed against potential delays to publication. (MEMR13: highly desirable) • **Consider**: Thorough documentation of the search process (dates, strings, filters, etc.) and its results (database exports of ALL screened results) combined with effective deduplication considerably reduces the workload of search updates.		• Is the literature sample up to date? • Is a re-run of the searches needed?

Box 2
Comparing STEP 2 to current keyword searching practice in management SLRs and MAs.

The review of 404 management SLRs and MAs revealed that many authors already, to some extent, concur with the sampling guidance we put forward. Almost all SLRs and MAs (96%) relied on database searches as their primary source, or one of their primary sources, to identify literature. Only 16% primarily used a journal-driven approach, and 5% used a seminal-works driven approach. Compared to Hiebl (2023), who found that 70% of the reviews in Academy of Management Annals and International Journal of Management Reviews between 2004 and 2018 used database-driven approaches, our results may indicate a higher relevance of database-driven approaches since then.

Review authors searched on average 4.12 databases (median=3), suggesting increasing rigor compared to previous years (Hiebl, 2023). The 393 studies that reported the databases and systems they searched used 147 different ones, with the most prevalent being Web of Science (213×), Scopus (148×) and APA PsycINFO (129×), GS (124×), and ScienceDirect (72×). Notably, nearly all studies (99%) used keyword searching, and about two-thirds (63%) searched at least three databases. This aligns with guidance from medicine and other fields that regard keyword search as the foundation of an SLR (Cooper et al., 2017; Frants et al., 1999; Higgins et al., 2023; Lowe et al., 2017). Upon closer inspection, however, we see all sorts of sampling deficiencies. For example, in one case, authors searched many databases with inadequate search strings, while another employed a suitable search string at a few inappropriate databases. Further, we identified that 11 studies (3%) relied exclusively on GS to inform their keyword searches, and some 38 studies (10%) used it with one or two other databases. Based on our and previous analysis (Gusenbauer and Haddaway, 2020), GS is not recommended as a principal database for keyword searches.

Keyword searches identified over 12 times more documents than other heuristics but only included six times more relevant documents (see Table 3). This illustrates the advantages of keyword searches over other heuristics in terms of sensitivity, yet also reveals its limitations in terms of precision. As a consequence, review authors should spend a great deal of time on crafting their keyword search strategy to achieve highly sensitive searches with acceptable levels of imprecision. Due to management literature's diverse language, composing comprehensive search

strings is critical to identifying all or at least almost all relevant literature on a topic.

Management reviews frequently underestimate the importance of crafting such comprehensive keyword search strategies, with lower median (median=20 keywords, mean=36, min=2, max=294) search string length than medical SLRs (median=50) (Bramer et al., 2018). It might surprise review authors, particularly those who only search a few databases with a few keywords, that experienced expert searchers and specialist librarians spend an average of 12.5 hours on this task (median=8). Less experienced searchers spend even longer (Bullers et al., 2018). These findings illustrate the diligence exhibited in medical SLRs, which work more thoroughly to circumvent the pitfalls of search strategy design (Salvador-Oliván et al., 2019).

To reduce the screening of irrelevant studies, authors must effectively select suitable databases and create precise keyword search strategies using Boolean operators. Most keyword searches employed "OR" and "AND" operators (87% and 65%), used quotations (73%), truncation (62%), and parentheses (54%), while only 7% used "NOT" operators, wildcards (6%), or proximity operators (3%) (see Fig. 2). While we did not review individual search strategies, we expect the anecdotal search string errors we identified to be common (Salvador-Oliván et al., 2019). From the quantitative results, we see that proximity operators are particularly underutilized. These operators may save considerable review time by increasing search precision by linking concepts more closely, ensuring search terms are collocated rather than scattered throughout the text.

Results of quality check #1 (keyword search strategy validity) and #2 (keyword search breadth validity)

While the two quality checks we propose are not yet reported, we still tried to gather evidence on how the 404 studies performed. First, to illustrate the usefulness of quality check #1 (keyword search strategy validity), we manually validated a randomly chosen SLR that identified documents via scoping searches (narrative searches and experts) and used keyword searches. We employed the same search strategy at one of the major databases the author reported. The resulting Scopus search revealed that the search strategy did not identify eight of all 29 documents the database covered from the review's sample. Thus, the check revealed that the search strategy had missed 28% of the relevant studies. These were most likely identified in the scoping searches, yet their language was not accounted for in the keyword search strategy. Thus, had the author performed this check, the search strategy would have been more sensitive, likely identifying many more relevant studies (details on how we compiled this check can be found in Supplementary Material 5).

Second, we found substantial differences in keyword search breadths (quality check #2), expressed in the varying number of documents that are identified and what portion of them is deemed relevant (NNR). Our findings show that management authors' keyword searches identified 4114.07 documents on average (including duplicates) (median=1364) with a maximum of 64,923 and a minimum of four. Review authors screened, on average, 1830.35 study titles/abstracts for relevance (median=808) in their keyword searches. This presents a substantial discrepancy: while one review screened only 22 studies, another screened 22,276. Accordingly, the screening yield varies, too. While some SLRs and MAs only screened about one study to identify a relevant document (NNR=1.125), others needed to screen as many as 221. Compared to medicine's best-practice levels (NNR ranging from 8 for exceptionally well-targeted searches to 73 for the most comprehensive (Bramer et al., 2017)), most management reviews screened considerably fewer studies, with a median NNR of 9.90 (mean 27.02). While most reviews screened up to a dozen documents to identify one relevant one, one management SLR even screened 7578 for every relevant study (it was not counted in the table as it did not report the keyword search in detail). So, what do these differences mean? Are review authors for which most identified documents are relevant just better searchers than those who screen hundreds of studies?

The answer is no. In most cases, these wide discrepancies in NNRs indicate a significant variance in search diligence and review scope. Overall, we identified that 42% of reviews screened fewer than eight documents to identify a relevant one (NNR<8), an indicator that many relevant studies may have been left unidentified in these reviews through overly narrow search areas. Upon reviewing the study with the lowest NNR (1.125), we indeed found that its keyword search strategy was profoundly inadequate, relying on overly limiting keywords that missed many relevant studies. Given social science's exceptionally difficult language, one would expect management researchers to be more diligent searchers, as reflected in higher NNRs. Unfortunately, the opposite is the case.

Conversely, about 6% of reviews screened more than 100 documents to identify a single relevant one (NNR>100)—an indicator that testifies to exceptional sampling efforts but can also be a sign of search ineffectiveness. We assessed these studies' overall keyword search strategies as diligent, yet potentially more effective had their breadth had been further narrowed somewhat. Additionally, we analyzed the search strategy of the SLR with the exceptionally high NNR of 7578 and found that the authors likely could have saved more than 80% of screening time simply by using a proximity operator in their search string—an issue this guide would have helped with. These freed up resources can be put to better use in additional backward or forward citation searches, which many studies do not yet pursue, or in the analysis, synthesis, and reporting parts of the reviews. Overall, reviews with very low NNRs may require guidance on how to search more thoroughly to identify more relevant studies, while reviews with very high NNRs require guidance on search effectiveness.

On average, keyword-based search and screening efforts in management SLRs and MAs yielded 75 documents (median=66) included in the final sample, with a maximum of 318 documents and a minimum of 11 (for a detailed overview, see Table 3).

Table 3
Searching and screening in management systematic literature reviews and meta-analyses from 2021.

	Detailed description of metric	Reported in X of 404 studies	Min	Max	Mean	Median
Keyword search	No. of databases	394 (97.28 %)	1	66	4.12	3
	No. of terms*	143	2	294	35.74	20
	No. of OR*	166	0	368	25.12	11
	No. of AND*	124	0	53	3.21	1
	No. of NOT*	14	0	2	0.10	0
	No. of studies identified (incl. duplicates)	260 (64.36 %)	4	64,923	4114.07	1364
Screening (keyword search only)	No. of studies identified (with single search string, incl. duplicates)	115 (28.47 %)	34	36,762	2325.31	886
	No. of studies screened (abstract, title, keywords)	133 (32.92 %)	22	22,276	1830.35	808
	No. of studies included (in final sample)	86 (21.29 %)	11	318	74.80	66
	No. needed to read (NNR)	**82 (20.30 %)**	**1**	**211**	**27.02**	**9.90**

(*continued on next page*)

Table 3 (*continued*)

	Detailed description of metric	Reported in X of 404 studies	Min	Max	Mean	Median
Screening (non-keyword search only)	No. of studies identified (incl. duplicates)	163 (40.35 %)	1	13,620	320.98	21
	No. of studies screened (abstract, title, keywords)	39 (9.65 %)	1	1048	111.38	32
	No. of studies included (in final sample)	81 (20.05 %)	1	69	12.10	7
	No. needed to read (NNR)	**14 (3.47 %)**	**1**	**44**	**6.07**	**1.00**
Screening (all search methods combined)	No. of studies identified (incl. duplicates)	350 (86.63 %)	22	736,860	6293.20	1367
	No. of studies screened (abstract, title, keywords)	334 (82.67 %)	22	712,365	4447.50	844
	No. of studies screened (in full text)	251 (62.13 %)	15	23,214	447.35	167
	No. of studies included (in final sample)	404 (100 %)	7	4488	118.20	73
	No. needed to read (NNR)	**333 (82.43 %)**	**1**	**7578**	**58.09**	**11.57**

* In case of multiple search strings, only the longest was considered.

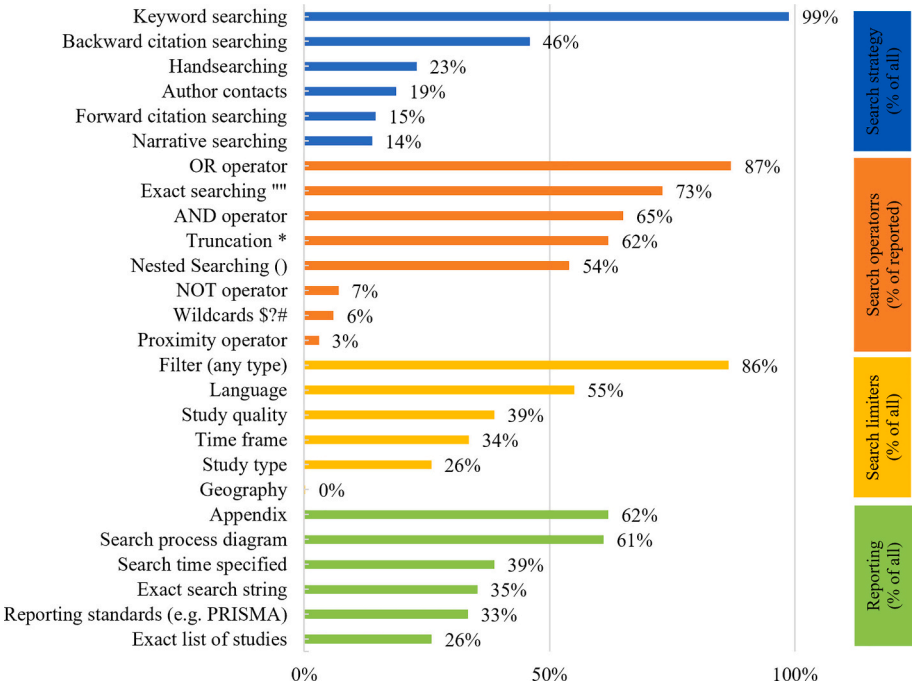

Fig. 2. Sampling in management systematic literature reviews and meta-analyses from 2021.

4.3. STEP 3: Backward and forward citation searching (for details, see Table 4)

Keyword searching, backward citation, and forward citation searching constitute the three *"major sources to find literature"* (Xiao and Watson, 2019)—Particularly in management, where *"a range of search techniques are required to comprehensively identify social science literature"* (Papaioannou et al., 2010, p. 121). While a keyword search identifies studies based on commonalities of language, a citation-based search identifies studies based on commonalities of topics. Such commonalities are judgments from authors who consider another study relevant to theirs. These non-language commonalities are essential for systematic search topics that are difficult to search via keywords—Instances where backward and forward citation searching should be considered *mandatory* (Briscoe et al., 2020, p. 170; Hirt et al., 2024).

4.3.1. What are "difficult-to-search" review topics?

Review topics may be difficult to search, despite keyword searching having been optimized (see STEP 2). An exact definition of topics that are difficult to search is impossible (Hirt et al., 2024). Hence, we attempt instead to enumerate identifying features of difficult-to-search topics, upon which authors may judge the appropriateness of citation searching.

4.3.1.1. Unclear concepts. Poor conceptual clarity is noticeable in topics with inconsistent terminology, vocabulary overlaps with unrelated topics, or historical development of terminology (Briscoe et al., 2020; C Lefebvre et al., 2023). Keyword searches typically use a combination of concepts that jointly delineate the topic and span the search area. To manage screening workload, management authors may omit specific ambiguous or fuzzy terms that would identify many irrelevant studies. In such cases, citation searching may help to identify additional relevant studies and terminology that was difficult to learn during scoping.

4.3.1.2. Low-precision keyword searches. One potential sign of difficult language is when review authors need to read many irrelevant studies

before they identify relevant ones (high NNR) (see quality check #2). For high-NNR keyword searches, citation searches will likely find additional relevant documents.

4.3.1.3. Low-yield searches. When keyword searches identify few relevant studies, citation searches will likely be effective, while the workload remains moderate due to the limited number of seed references.

4.3.1.4. Difficult-to-access language (Linder et al., 2015; Papaioannou et al., 2010). If the language describing the targeted concepts does not appear in title, abstract, or keywords, search strings will have difficulty in identifying the documents. This is especially true when authors of MAs attempt to identify studies with particular control variables. Quality check #1 highlights relevant documents identified by narrative searches whose title, abstract, keyword information contains few topic-related keywords and thus make these documents difficult to identify. If so, authors may still attempt a full-text search in databases that support it (see Search Smart filter: *"Query > Full-text search"*). Nevertheless, in most cases, databases only have a very limited number of references with associated full-text information (e.g., OpenAlex, Lens, or ABI/Inform Global) or provide very limited search capabilities (e.g., GS). In the absence of good full-text retrieval capabilities, citation searches are likely a helpful technique to identify additional documents with difficult-to-access language.

4.3.2. What are "easy-to-search" review topics?

As citation searching *"represents a significant additional investment of searching effort"* (Wright et al., 2014, p. 2), we also want to list instances when it is *"not explicitly recommended."* Such easy-to-search topics are characterized by a sensitive search (Hirt et al., 2024). Thus, if authors have already identified a large number of relevant studies—particularly when keyword searches seem effective, as evidenced by low NNRs—the additional workload of citation searches offers a lesser payoff. Keyword searches will likely be most effective when language is specific and explicitly described in titles/abstracts/keywords of documents, and concepts can be clearly defined with PICOC elements (Hirt et al., 2024). For example, the present study searched for systematic reviews and meta-analyses—a very specific topic to which study authors prominently refer in documents' titles. In such cases, supplementary citation searching adds little value and thus is not explicitly recommended.

4.3.3. Backward or forward citation searching?

The TARCiS statement recommends employing both backward and forward citation searching for difficult-to-search topics (Hirt et al., 2024). This guidance is particularly relevant because backward citation searching and forward citation searching each serve unique purposes: backward citation searching is instrumental in uncovering the foundational studies of a topic, while forward citation searching is crucial for capturing the most recent research or evolving fields where terminologies may have shifted. Once researchers have pinpointed the foundational articles, such as those that describe key methods, concepts, or theories, forward citation searching can be used to identify subsequent studies that are conceptually or methodologically linked to these articles—regardless of whether they share the same language. SLRs and MAs with difficult-to-search topics that wish to focus on just one would need to provide good arguments for why either backward or forward citation searching alone would yield adequate results.

4.3.4. Employing backward citation searching

Backward citation searching via database downloads instead of manual reference list checking is preferable, as it allows authors to screen the deduplicated results based on title/abstract information and to easily keep track of the screening process. Backward citation searching seeks references cited in a sample of relevant seed references that authors need to determine first. These seed references are composed of all identified relevant documents from STEPS 1 and 2. According to our analysis of 108 documents, review authors may expect on average about 71 backward citations for each seed reference—a number that is increasing every year as reference lists grow ever longer (Sánchez-Gil et al., 2018).

To find the most suitable backward citation databases that also cover most management documents, we use Search Smart and select the *"Business, Management and Accounting"* subject and the *"Backward citation searching"* preset including *"Bulk select & export options."* To obtain a sorted list with management databases with the most accurate backward citation information at the top, we sort the databases with *"Backward cit. score x Subj. cov. (asc.)."* We find Lens as the highest ranked option due to its superior business coverage and good backward citation information. Second-best is WOSCC (preferable over Scopus), which has somewhat better backward citation information yet inferior management document coverage.

While researchers often rely on a single database for backward citation searching, recent advice advocates searching more than one (Hirt et al., 2024), as individual databases' citation information is rarely complete.

4.3.5. Employing forward citation searching

Forward citation searching is used by authors to identify studies that cite a relevant document. As citations accumulate over time, the method yields more results the older the references in the seed sample.

Search Smart also indicates the most suitable databases for forward citation searching. Employing similar settings as for backward citation searching but using the *"Forward citation searching"* preset instead, we

Table 4
Step 3—backward and forward citation searching (mandatory for difficult-to-search topics).

Procedural details	Tools/help	Evaluation questions
Necessity of citation searching • **Backward and forward citation searching is mandatory for difficult-to-search topics** (criteria on how to determine difficult-to-search topics are listed in the previous section). For easy-to-search topics, citation searching is "not explicitly recommended" (MEMR6: mandatory; MEMR19: mandatory). • Citation searching must always build on preceding comprehensive keyword searches (STEP 2) (MEMR25: mandatory).	• If you struggle to assess whether your topic is difficult to search, *ask a specialized librarian or similar information specialist for help.*	• Is the (non-) relevancy of citation searching well justified?
Seed references • After documents identified in STEPs 1 and 2 have been screened for eligibility, all relevant documents **("seed references")** should be used for backward and forward citation searching (MEMR20: mandatory).	• *Reference management software* will help to keep track of seed references.	• Are all studies identified as relevant through keyword searches considered as seed references?

(*continued on next page*)

Table 4 (*continued*)

Procedural details	Tools/help	Evaluation questions
• Particularly for small seed samples, you may consider additionally examining the (backward and forward) citations of documents that were only marginally excluded in the screening process.		
Backward citation searching • **We recommend backward citation searching a suitable database over manual reference list checking, for elevated screening efficiency and accuracy.** Doing so permits the documents' deduplication and gives authors title/abstract or full-text information (MEMR23: highly desirable). Consequently, the reviewer has to screen fewer documents and also has additional information with which to assess eligibility (compared to screening decisions based on titles only) (MEMR21: highly desirable). • **The best database for backward citation searching is Lens, followed by WOSCC with full coverage of the underlying databases (over Scopus).** • The best open alternatives for backward citation searching are **Lens** and **Semantic Scholar**. However, they are relatively inaccurate, missing backward citations or listing references that were not cited. • Researchers without access to the most recommended databases are advised to review reference lists of full texts directly ("reference list checking"). Reference lists of the original publishers, such as *Sage Journals Online* and *ScienceDirect*, also tend to be accurate. However, identified documents will be available at different publishers, making it more labor-intensive to find these documents. • Backward citation searching (or reference list checking) may be used to validate the sensitivity of the keyword search (Hirt et al., 2024) (MEMR21: highly desirable).	• *Citation Chaser* software (Haddaway et al., 2022a): download all (deduplicated) backward citations covered by *Lens*. • *Search Smart* (Gusenbauer, 2024b): to identify (additional) suitable databases with the most powerful citation indices.	• Is backward citation searching (alternatively: reference list checking) employed? • Is a comprehensive database selected for backward citation searching?
Forward citation searching • Citation indices are all different (Martín-Martín et al., 2021). Choosing smaller or larger forward citation indices can help focus search results: If your goal is to include a broad range of document types and qualities (incl. Gray literature, conference articles, dissertations, etc.), then **use GS** (alternative: *Lens*). If the goal is to focus on high-quality journal publications, **use Scopus** (alternative: *WOSCC*, but only if your institution has a comprehensive subscription). • **Scopus** or **WOSCC** only retrieve about 57–66 % of the forward citations of **GS**, whose citation index is the gold standard across databases (Gusenbauer, 2024a).	• *Publish or Perish* (Harzing, 2007): download forward citations of a single document on *GS* (*"Retrieve citing works"*) and combine them manually. • *Citation Chaser:* download all (deduplicated) forward citations covered by *Lens*. • *Search Smart* (Gusenbauer, 2024b): to identify (additional) suitable databases with the most powerful citation indices.	• Is forward citation searching employed? • Is a comprehensive database selected for forward citation searching?
Multiple citation indices • Searching two (instead of one) suitable indices for backward citation searching and two for forward citation searching is recommended, as it will increase recall (MEMR22: highly desirable).	• *Search Smart* (Gusenbauer, 2024b): to identify (additional) suitable databases with the most powerful citation indices.	• Are multiple citation indices searched?
Deduplication • If review authors screen citations *after* all backward or forward citations have been downloaded and deduplicated, they may expect to save about 25–27 % (backward citations) and 20–23 % (forward citations) of their screening time (own analysis) (MEMR23: highly desirable). • Expect additional screening time savings if: - documents are retrieved from multiple citation indices (they will overlap extensively). - documents can be deduplicated against the already-screened studies from STEPS 1 and 2. - a portion of documents falls outside the review scope (see eligibility criteria) and you can employ filters, e.g., documents with excluded time frames or document types.	For deduplication suggestions, see STEP 1	
Iterating searches • If citation searches find additional eligible documents, another iteration of citation searching must be considered using these records as new seed references. • If significant new keywords emerge, keyword searching must be repeated including these keywords (MEMR24: mandatory) (see STEP 2).		• Upon identification of significant new documents, are citation searches updated? • Upon identification of significant new terminology, is the keyword search updated?

Box 3
Comparing STEP 3 to current citation searching practice in management SLRs and MAs.

The general importance of backward citation searching is reflected by its prevalence in medical SLRs (Higgins et al., 2023). While 81% of Cochrane SLRs use backward citation searching (Briscoe et al., 2020), only 46% of SLRs and MAs in management do so. Management SLRs and MAs, at 15%, perform forward citation searches as infrequently as Cochrane SLRs (12%) (Briscoe et al., 2020). The less frequent application of forward citation searching in medicine is likely attributable to MECIR guidance that recommends yet does not mandate it. Given that management topics are presumably more difficult to search than those in medicine, employing multiple search techniques is likely to lead to considerably better sampling outcomes.

Our analysis found that non-keyword-based search strategies (predominantly citation searches) are essential to comprehensive searching in management. In the 39 studies that reported both techniques, non-keyword searches discovered 18% of the unique documents that were missed by keyword-based searches. The relevance of non-keyword-based searches would be even greater if one counted the relevant references they jointly identified with keyword searching, and if these studies had implemented all the non-keyword-based searches recommended for their study topic (see Table 3).

In 2021, only about 9% of the SLRs and MAs in management conducted all three steps: keyword searching in at least three databases followed by a backward and a forward citation search. Two examples that use these search steps diligently are Ryu et al. (2020) and Boness et al. (2021), both meta-analyses published in psychology-related journals; however, they should not be considered as role models.

again find that Lens, followed by Scilit, and Scopus, cover the most forward citations and also allow bulk downloads. Alternatively, GS's citation index would be the most comprehensive index (Gusenbauer, 2024a), yet it does not natively support bulk downloads and includes a few citations to nonexistent works (Ibrahim et al., 2024). If authors wish to use GS for its more complete forward citations of varying (and often inferior) quality (Wright et al., 2014), they can use Publish or Perish (Harzing, 2007), a software application that allows the search and retrieval of up to 1000 GS citations per publication year for individual documents. If the SLR exclusively focuses on higher-quality (journal) publications, review authors should search Scopus (over WOSCC) for more tailored search results.

4.4. STEP 4: Reporting the literature sampling process (for details, see Table 5)

"Readers should not have to infer what was probably done; they should be told explicitly" (Altman, 1996, p. 570). Adequate reporting of sampling is a necessary condition for well-crafted SLRs and MAs. Reporting sampling decisions and results enables readers and (peer) reviewers to assess the review's sampling quality and also allows for better replication of the results.

4.4.1. Reporting expectations in management reviews

Reporting requires a detailed description of the searching and screening process so it can be replicated without additional information. Reporting should include a transparent description of how and when which databases were searched, the number of relevant studies identified, and why documents were excluded. Reporting should explain how searches were run, including the entire search string, field codes, and applied filters. Reporting the screening process includes describing all screening steps and the decisions taken until the final sample of relevant studies was identified. Table 5 outlines the Reporting Expectations in Management Reviews (REMR1–19) adapted and synthesized from established guidelines (for details, see Supplementary Material 4). Each REMR item is linked to the corresponding sampling step. REMR16 and REMR17 are new items associated with the quality checks (for details, see Table 5) included in this guide. Reporting these two quality checks ensures readers get a better picture of keyword search diligence.

4.4.2. Adequate reporting supports the cumulativeness of science

If items are reported well, future reviews can build on search and screening decisions from previous reviews by reusing search strategies and reference lists. Reusing well-conceived search strings improves review quality and effectiveness and is common practice in the medical disciplines. For example, Terwee et al. (2009) diligently constructed a search string that has helped hundreds of studies to identify studies using *measurement of patient-reported outcomes*. While management research has not yet established such practices of cumulative learning in management reviews, the Center for Evidence-Based Management (CEBMa) is a noteworthy example publishing reusable search strings (CEBMa, 2022).

4.4.3. Using the appendix for comprehensive reporting

As space is limited in review studies, authors should use the appendix or supplementary-materials sections to include all additional necessary material to make the SLR's process and outcomes transparent and reproducible. Withholding essential information and replacing it with a "*data available upon request*" statement is not good practice, as a study (Gabelica et al., 2022) found that authors of 93 % of the studies reviewed were non-responsive or unwilling to share data, despite indicating otherwise in their data-sharing statements.

Review authors who have finalized scoping, searching, screening, and reporting may proceed with the next steps of the review: quality appraisal, data extraction, and data analysis, which determine the quality of the knowledge synthesis (see Fig. 1). It is important to note that other aspects of the review need to be reported too, for which PRISMA and other reporting guidelines also offer recommendations.

Table 5
Step 4—Reporting the literature sampling process (mandatory).

Procedural details	Tools/help	Evaluation questions
Reporting the literature sampling process • **Reporting requires a detailed description of the searching and screening process so it can be replicated without additional information** (MEMR11/MEMR16: mandatory)**. Use the PRISMA flow diagram and the reporting items listed below.** • **The review topic and scope choices, including those that limit review comprehensiveness, must be transparently reported from the outset** (title, abstract) so readers may judge the review's scope, representativeness, and validity. Narrow understandings of concepts, limited timeframes, or a limited selection of journals are significant scope limitations. • **Use standardized language** to refer to "forward/backward citation searching," or "reference list checking," and "seed references" throughout the review (MEMR18: highly desirable).		• Are important scope limitations transparently reported in the title or abstract?
Reporting using the PRISMA flow diagram • **The PRISMA 2020 flow diagram for new (updated) systematic reviews, which included searches of databases, registers, and other sources (Page et al., 2021), should be used to structure the reporting of the searching and screening process for STEPS 1–3** (MEMR26: mandatory) • In PRISMA, the keyword search (STEP 2) should be reported under *"Identification of studies* via *databases and registers."* • In PRISMA, backward and forward citation searches (STEP 3) and the scoping searches (STEP 1) consisting of narrative searches, handsearches, author contacts, etc. should be reported under *"Identification of studies* via *other methods."*	• *PRISMA2020,* an R package and Shiny app, helps create PRISMA 2020 flow diagrams (Haddaway et al., 2022b).	• Is a comprehensive flow diagram (preferably PRISMA) used to report the sampling process?
Reporting items (and their corresponding steps, details see Supplementary Material 4) • REMR1 (for STEP 1): If applicable, provide registration information for the review, including register name and registration number. Indicate where the review protocol can be accessed, or state that no protocol was prepared. Describe and explain any amendments to information provided at registration or in the protocol. • REMR2 (for STEP 1): Specify the inclusion and exclusion criteria for the review. Describe changes to prespecified inclusion criteria and when they were made *(also reported in PRISMA flow diagram)*. • REMR3 (for STEP 1): Describe any online or print source purposefully searched or browsed (e.g., previous reviews, related reviews, specific journals and volumes) and how this was done. • REMR4 (for STEP 1): Indicate whether additional studies or data were sought by contacting authors, experts, manufacturers, or other sources *(also reported in PRISMA flow diagram)*. • REMR5 (for STEPS 2–3): Name each individual database used for various search types (keyword searches, backward citation searches, forward citation searches, etc.). State the coverage (retrospective or database suite) and the platform (i.e., search system) for each database. If databases were searched simultaneously on a single platform, state the name of the platform, listing all of the databases searched. If applicable, state the tools that were used to access them. Specify the date when each source was last searched *(also reported in PRISMA flow diagram)*. • REMR6 (for STEP 2): Include the search strategies for each database and information source, copied and pasted exactly as run. • REMR7 (for STEPS 2–3): Specify that no limits were used, or describe any limits or restrictions applied to a search (e.g., date or time period, language, study design) and provide justification for their use. • REMR8 (for STEP 2): Indicate when search strategies from other literature reviews were adapted or reused for a substantive part or all of the search, citing the previous review(s). • REMR9 (for STEP 2): Describe the search peer review process, which can be conducted by an experienced searcher, information specialist, or librarian, and/or by following the "PRESS 2015 Evidence-Based Checklist." • REMR10 (for STEP 3): Report which seed references were used for citation searching (along with a justification should the seed references differ from the set of included records from the results of the primary database search). • REMR11 (for STEP 3): If applicable, state the number of citation searching iterations (and possibly the reason for stopping, if the last iteration still retrieved additional eligible records). • REMR12 (for STEPS 1–3): Describe any additional information sources or search methods used. • REMR13 (for STEPS 1–3): Document the total number of records identified from each database and each search method (e.g., keyword searching, backward and forward citation searching, handsearching, narrative searches) and other information sources *(also reported in PRISMA flow diagram)*. • REMR14 (for STEPS 1–3): Describe the processes and any software used to deduplicate records from multiple database searches and other information sources *(also reported in PRISMA flow diagram)*. • REMR15 (for STEPS 1–3): Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many authors screened each study (based on title/abstract information and later full text information), their experience level, whether	• A solution for extensive reporting despite limited journal space is to use the *appendix/supplementary materials section* to present parts of the information.	• Are all steps of the searching and screening process transparently reported (REMR1–19)?

(*continued on next page*)

Table 5 (*continued*)

Procedural details	Tools/help	Evaluation questions
they worked independently. Describe how disagreements were resolved if multiple authors were used. Describe how multiple studies were handled and the handling of studies that did not contain sufficient information to judge eligibility. State whether studies identified in STEPS 1–3 were screened the same way, or if not, describe the alternative method used. If applicable, provide details regarding the types of automation tools used, including any AI tools, and describe their role in the sampling process *(also reported in PRISMA flow diagram)*. • REMR16 (for quality check #1 in STEP 2): To report keyword search strategy validity, report the number of relevant documents identified through scoping in STEP 1 (handsearches, narrative searches, etc.) and the share of these documents that was identified by the main keyword search strategy from STEP 2. As not all relevant documents will be available on that database, calculate the share only for those covered by the database (details see Table 2, example see Supplementary Material 5). • REMR17 (for quality check #2 in STEP 2): To report keyword search breadth validity, report the total number needed to read (NNR) from the keyword searches conducted in STEP 2. Do this by dividing the number of identified documents via keyword searches (only those that were actually screened, i.e. after deduplication, etc.) by the number of relevant ones within the identified set (details see Table 2). • REMR18 (for STEPS 1–3): Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram. Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded *(also reported in PRISMA flow diagram)*. • REMR19 (for STEPS 1–3): Cite each included study and present its characteristics.		

Box 4
Comparing STEP 4 to current reporting practice in management SLRs and MAs.

As observed earlier (Simsek et al., 2021), there is substantial room for improvement in the reporting of sampling. While many studies (61%) provide some diagrammatic illustration of their search process—albeit often confusing or limited—only about half of those (33%) followed established guidance (e.g., PRISMA). This reveals a gap between the authors' intention to report SLRs transparently and their understanding of the information required for transparency and replicability. Accordingly, the importance of conduct and reporting guidelines cannot be overstated. Readers seem to value thorough review work, as transparency through better reporting has been found to correlate with research impact (Simsek et al., 2021). Overall, the quality of reporting would improve significantly if such guidance was universally applied.

In the absence of guidance on which items ought to be reported, studies seem to report only a fraction of them. Most notably, only 143 studies (35%) specify the exact search string used in their keyword searches. A majority of studies (61%) do not adequately report the database(s) searched, as they either do not report or only partially report database searches, or report searches via platforms (ProQuest, EBSCOhost, WOS, or the WOSCC) without stating the specific databases or indices (e.g., SCI, SSCI, Business Source Premier), leaving out critical information for reproducing or evaluating their search strategies. For example, reporting the WOSCC is not sufficient, as it is not a standardized package but differs in the databases included and the retrospective coverage they offer.

Only 24% of studies specified the field codes they utilized—crucial information that states which parts of a document were searched (e.g., title, abstract, and keywords). Further, only 39% specified the time period the literature search covers, while only 4% provided the exact search date when individual databases were searched. As database coverage increases continuously and functionality changes periodically, this information is relevant to assess the search procedure.

Our review shows that only 26% of management SLRs and MAs report the complete list of studies. Since few review authors respond to data requests (Gabelica et al., 2022), information on the review sample is rarely obtainable. Even more important is the usage of appendices or supplementary materials to report such information, which 62% of studies included. These reporting issues make it impossible for peer reviewers and readers to replicate the search process and assess the review's diligence and scope.

5. Discussion

This study presents a detailed, four-step guide on literature sampling when conducting SLRs and MAs in management and adjacent social science disciplines. The guide's recommendations respond to persistent issues in sample selection as part of SLRs in management (Kunisch et al., 2018) and calls to improve sample selection from other disciplines (Higgins et al., 2023; Simsek et al., 2021).

The guide's recommendations are based on two pillars. First, to learn *"how management researchers search,"* we reviewed the search, screening, and reporting process of 404 SLRs and MAs published in journals rated 3, 4, and 4* by AGJ in 2021 (see Fig. 2 and Table 3) and compared it to the recommendations of this guide (see Boxes 1-4 in the sections on STEPS 1–4). To understand the source of sampling deficiencies, we reviewed the suggestions of all 135 documents that review authors used as sampling guidance. Second, to learn *"how to search,"* we adapted the best guidance on SLR conduct and reporting to the management field (see Fig. 1, Tables 1–2 and 4–5, and Supplementary Materials 3 and 4). To learn *"where to search,"* we collected information on the coverage and functionality of 50 bibliographic databases of particular relevance to management (see Supplementary Material 1) and made them available at Search Smart, where review authors can identify the best databases for their search needs (and can easily adapt this guide's recommendations to other disciplines).

5.1. How to apply this guide

The four-step guide (see Tables 1–2 and 4–5 for details) aims to help SLR and MA authors make best use of their limited resources for comprehensive literature sampling. Effectiveness and efficiency are achieved through the recommendation of selected but highly relevant search heuristics and databases. Following the guide gives authors a solid minimum basis that they may choose to go beyond, particularly if search yield is still low.

The guide was designed to be generally applicable and yet versatile enough to account for differences in review topics. Authors of SLRs and MAs are invited to adapt the procedures within the workflow's four steps to reflect the unique scope and context of their review. For example, broad, fuzzy concepts will likely require particularly careful scoping (STEP 1). Finding the right keywords may be difficult, making systematic keyword searching (STEP 2) difficult too, requiring sophisticated search strings that make it particularly challenging to balance search sensitivity with precision. Authors may need to search more databases than the suggested minimum number. Moreover, for such hard-to-search topics, the associative search heuristics of backward and forward citation searching (STEP 3) are particularly important. Conversely, SLRs on clear and distinctive concepts will find it easier to identify literature based on keyword matches.

The most effective way to apply the individual steps will also depend on the specifics of the SLR. Thus, within each of the steps, the guide details important decisions and procedures that authors must consider when determining their specific search strategy (see Tables 1–2 and 4–5). While some procedures are mandatory (expected to be applied; justification needed if not), others are highly desirable (should generally be applied, but no justification needed if not) or optional (may be applied or not). For example, during scoping, authors are free to determine suitable eligibility criteria, yet they must follow mandatory procedures to explore the language and previous research within this scope. For many other decision points, the guide documents individual reasons for how authors may, should, or must proceed.

As the authors of this guide, we acknowledge that we too are continuously learning how to sample most productively and efficiently. Accordingly, we refrain from retrospectively pointing the finger at studies that could have been better. We believe in a constant ambition to unearth new insights into review and sampling techniques so we can conduct more thorough, relevant, and thus also impactful SLRs and MAs in the future.

5.2. Outlook on the use of AI in systematic searching

With the release of ChatGPT in late 2022 and its significant impact on education and research (Bommarito and Katz, 2023; Else, 2023; Grimaldi and Ehrler, 2023; Susnjak, 2022), researchers and information specialists are wondering about the qualities and limitations of LLMs in systematic review work (Hutson, 2022). On the one side, LLMs seem to support document discovery (Hutson, 2022), search string creation (Wang et al., 2023), or screening work (Issaiy et al., 2024). On the other side, LLMs seem to hallucinate, be confident but wrong (Galactica, 2022; Heaven, 2022; Hosseini et al., 2023), or make hard-to-detect errors (Else, 2023).

At a time when LLM-based tools are evolving rapidly, and there are few comprehensive assessments of how well they already adhere to SLR quality standards (Gusenbauer, 2023), we must be cautious in using them in SLRs and MAs. Otherwise, we will risk introducing new types of biases. While there is no clear guidance regarding the use of LLM-based tools as yet (Parisi and Sutton, 2024), we still can make suggestions knowing which parts of SLRs are less systematic. Thus, we can recommend LLM-based tools in exploratory scoping searches (STEP 1), where AI may help address the problem of concepts being described in language that is difficult to know in its entirety (Boell and Cecez-Kecmanovic, 2015). With vector-based search capabilities, AI can help elevate sampling strategies beyond keyword matching and associative search strategies. However, due to LLM's irreproducible and non-transparent architecture, LLM-assisted literature discovery tools must still be treated like other functionally inadequate systems—that is, only to supplement (Hope et al., 2022) and never replace principal search strategies (STEPS 2–3).

5.3. Limitations

This guide describes minimum standard for sampling in SLRs, MAs, and other types of reviews that seek comprehensiveness, reproducibility, and transparency. While the guide will support good sampling quality, its recommendations should be considered in the light of specific limitations.

First, while rigorous sampling is a necessary condition for high-quality systematic reviews and is significantly related to review impact (Simsek et al., 2021), it should not distract from other, equally important aspects of reviews (Hiebl, 2023). While authors need to show diligence in sampling, this should not devolve into mere virtue signaling to deflect attention from weaker parts of the review. Sampling diligence should not lead to authors *"[…] pay[ing] more attention to the mechanics of review rather than the content of sources of material collected"* (Adams et al., 2017, p. 447).

Second, the guide's database recommendations are based on the latest available information provided by *Search Smart* (Gusenbauer, 2024b) on almost all of the databases management reviewers currently use. These recommendations depend on the quality of the underlying testing methods—that is, Query Hit Counts (Gusenbauer, 2019), backward and forward citation scores (Gusenbauer, 2024a), Basket of Keywords (Gusenbauer, 2022), metamorphic testing, and interface reviews (Gusenbauer and Haddaway, 2020). These tests yield the currently most accurate database recommendations for systematic searching (STEP 2) and backward and forward citation searching (STEP 3). Nevertheless, it is important to note that these database recommendations are based on estimates using the coarse 26-category All Science Journal Classification (ASJC) (Elsevier, 2024). Thus, review topics on specific sub-disciplines may need to search databases not covered in the analysis.

Third, the guide's recommendations will require updates in the future due to changes in databases' performance, new databases being introduced, or developments in AI-enabled sampling. These new developments will be captured either by updates of Search Smart's recommendations or by updates to the sampling recommendations of the sources this guide relies on. Specifically, Cochrane is currently investigating the formulation of rules on AI use in evidence synthesis (Cochrane, 2024). Thus, readers are advised to monitor updates in this space.

5.4. Conclusion

As we write, the influx of low-quality SLRs (and MAs) is undermining the method's reputation as the gold standard of knowledge creation and a cornerstone of theory-building. We believe recent criticism of SLR practice (Ioannidis, 2016; Moore et al., 2022) does not reflect an inherent problem of the method but an unwanted byproduct of its current success. More and more researchers are being lured into publishing SLRs and MAs by a fantasy of quick wins and high citation rates.

To assist review authors in achieving more methodical rigor and to respond to a lack of appropriate sampling guidance in management, we developed this four-step SLR guide to provide detailed practical recommendations. Effective sampling is mostly characterized by 1) a well-scoped review topic; 2) validated keyword searches representatively covering the phenomenon and characterized by a reasonable NNR, typically ranging between eight and 100; 3) keyword searches of a few highly suitable databases supplemented by 4) backward and forward citation searches for difficult-to-search topics; and 5) transparent reporting of every step. The specific recommendations of this guide are

designed not only to help review authors improve their sampling methods, but also to support peer reviewers and editors (see evaluation questions in Tables 1–2 and 4–5), collectively contributing to the unstinting rigor that readers of SLRs and MAs rightly expect. However, it is important to recognize that not all literature reviews need to be SLRs and that other types of literature reviews can also be contributive, valuable, and useful for the advancement of management research. Nonetheless, when reviews are designated as SLRs, they should adhere to the rigorous sampling practices established by the most respected sources outlined in this guide.

Disclosure statement

Michael Gusenbauer is the creator of Search Smart (https://www.searchsmart.org/), a non-proprietary website to compare academic databases and search systems. The authors received no financial support for the research, authorship, and/or publication of this article.

CRediT authorship contribution statement

Michael Gusenbauer: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sebastian P. Gauster:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization.

Acknowledgments

We would like to thank the editor and the two anonymous reviewers for their helpful guidance on developing this study. We are grateful to Prof. Denise Rousseau *(Carnegie Mellon University; Co-Chair, Business & Management Coordinating Group of Campbell Collaboration)*, Prof. Martin Hiebl *(Johannes Kepler University Linz),* Eric Barends *(Center for Evidence-based Management; Co-Chair, Business & Management Coordinating Group of Campbell Collaboration),* and Prof. Anne-Wil Harzing *(Middlesex University London)* for their valuable suggestions and comments on earlier versions of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.techfore.2024.123833.

Data availability

All data is shared in the Supplementary Materials.

References

Aagaard, T., Lund, H., Juhl, C., 2016. Optimizing literature search in systematic reviews - are MEDLINE, EMBASE and CENTRAL enough for identifying effect studies within the area of musculoskeletal disorders? BMC Med. Res. Methodol. 16 (1), 1–11. https://doi.org/10.1186/s12874-016-0264-6.

Adams, R.J., Smart, P., Huff, A.S., 2017. Shades of Grey: guidelines for working with the Grey literature in systematic reviews for management and organizational studies. Int. J. Manag. Rev. 19 (4), 432–454. https://doi.org/10.1111/ijmr.12102.

Aguinis, H., Cummings, C., Ramani, R.S., Cummings, T.G., 2020. "An A is an A": the new bottom line for valuing academic research. Acad. Manag. Perspect. 34 (1), 135–154. https://doi.org/10.5465/amp.2017.0193.

Aguinis, H., Ramani, R.S., Alabduljader, N., 2023. Best-practice recommendations for producers, evaluators, and users of methodological literature reviews. Organ. Res. Methods 26 (1), 46–76. https://doi.org/10.1177/1094428120943281.

Altman, D.G., 1996. Better reporting of randomised controlled trials: The CONSORT statement. In: Bmj, Vol. 313, Issue 7057. British Medical Journal Publishing Group, pp. 570–571.

Appelbaum, M., Cooper, H., Kline, R.B., Mayo-Wilson, E., Nezu, A.M., Rao, S.M., 2018. Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. Am. Psychol. 73 (1), 3–25.

Armstrong, R., Hall, B.J., Doyle, J., Waters, E., 2011. "Scoping the scope" of a cochrane review. J. Public Health 33 (1), 147–150. https://doi.org/10.1093/pubmed/fdr015.

Ayiku, L., Levay, P., Hudson, T., Finnegan, A., 2020. The NICE UK geographic search filters for MEDLINE and Embase (Ovid): Post-development study to further evaluate precision and number-needed-to-read when retrieving UK evidence. Res. Synth. Methods 11 (5), 669–677.

Bachmann, L.M., Coray, R., Estermann, P., Ter Riet, G., 2002. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. J. Am. Med. Inform. Assoc.: JAMIA 9 (6), 653–658. https://doi.org/10.1197/jamia.m1124.

Barends, E., Rousseau, D.M., 2018. Evidence-Based Management: How to Use Evidence to Make better Organizational Decisions, 1st ed. Kogan Page Publishers.

Barends, E., Rousseau, D.M., Briner, R.B., 2017. CEBMa Guideline for Rapid Evidence Assessments in Management and Organizations. Center for Evidence Based Management, Amsterdam.

Beyer, F., Wright, K., 2011. Comprehensive Searching for Systematic Reviews: A Comparison of Database Performance. Centre for Reviews and Dissemination, University of York, York.

Boell, S.K., Cecez-Kecmanovic, D., 2015. On being "systematic" in literature reviews in IS. J. Inf. Technol. 30 (2), 161–173. https://doi.org/10.1057/jit.2014.26.

Bommarito, M.J., Katz, D.M., 2023. GPT takes the Bar exam. SSRN Electron. J. 1–13. https://doi.org/10.2139/ssrn.4314839.

Boness, C.L., Watts, A.L., Moeller, K.N., Sher, K.J., 2021. The etiologic, theory-based, ontogenetic hierarchical framework of alcohol use disorder: A translational systematic review of reviews. Psychol. Bull. 147 (10), 1075.

Booth, A., 2008. Unpacking your literature search toolbox: on search styles and tactics. Health Inf. Libr. J. 25 (4), 313–317. https://doi.org/10.1111/j.1471-1842.2008.00825.x.

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.

Booth, A., Papaioannou, D., Sutton, A., 2016. Systematic Approaches to a Successful Literature Review, 2nd ed. Sage Publications Ltd.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2009. Introduction to Meta-Analysis. In: John Wiley & Sons. John Wiley & Sons.

Bramer, W.M., Rethlefsen, M.L., Kleijnen, J., Franco, O.H., 2017. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. Syst. Rev. 6 (1), 1–12.

Bramer, W.M., Rethlefsen, M.L., Mast, F., Kleijnen, J., 2018. Evaluation of a new method for librarian-mediated literature searches for systematic reviews. Res. Synth. Methods 9 (4), 510–520. https://doi.org/10.1002/jrsm.1279.

Briner, R.B., Denyer, D., Rousseau, D.M., 2009. Evidence-based management: concept cleanup time? Acad. Manag. Perspect. 23 (4), 19–32.

Briscoe, S., Bethel, A., Rogers, M., 2020. Conduct and reporting of citation searching in Cochrane systematic reviews: A cross-sectional study. Res. Synth. Methods 11 (2), 169–180.

Bullers, K., Howard, A.M., Hanson, A., Kearns, W.D., Orriola, J.J., Polo, R.L., Sakmar, K. A., 2018. It takes longer than you think: librarian time spent on systematic review tasks. J. Med. Libr. Assoc.: JMLA 106 (2), 198–207. https://doi.org/10.5195/jmla.2018.323.

Caulley, L., Catalá-López, F., Whelan, J., Khoury, M., Ferraro, J., Cheng, W., Husereau, D., Altman, D.G., Moher, D., 2020. Reporting guidelines of health research studies are frequently used inappropriately. J. Clin. Epidemiol. 122, 87–94. https://doi.org/10.1016/j.jclinepi.2020.03.006.

CEBMa, 2022. Center for Evidence-Based Management. https://cebma.org/resources-and-tools/methodological-search-filters/.

Chartered Association of Business Schools, 2021. Academic Journal Guide 2021: Methodology.

Chen, V.Z., Hitt, M.A., 2021. Knowledge synthesis for scientific management: practical integration for complexity versus scientific fragmentation for simplicity. J. Manag. Inq. 30 (2), 177–192.

Cochrane, 2024. Artificial intelligence technologies in cochrane. https://training.cochrane.org/resource/artificial-intelligence-technologies-in-cochrane.

Collins, A., Miller, J., Coughlin, D., Kirk, S., 2015. The Production of Quick Scoping Reviews and Rapid Evidence Assessments: A how to Guide. November, p. 59.

Cooper, C., Booth, A., Britten, N., Garside, R., 2017. A comparison of results of empirical studies of supplementary search techniques and recommendations in review methodology handbooks: a methodological review. Syst. Rev. 6 (1), 1–16. https://doi.org/10.1186/s13643-017-0625-1.

Cooper, C., Booth, A., Varley-Campbell, J., Britten, N., Garside, R., 2018a. Defining the process to literature searching in systematic reviews: A literature review of guidance and supporting studies. BMC Med. Res. Methodol. 18 (1), 1–14. https://doi.org/10.1186/s12874-018-0545-3.

Cooper, C., Varley-Campbell, J., Booth, A., Britten, N., Garside, R., 2018b. Systematic review identifies six metrics and one method for assessing literature search effectiveness but no consensus on appropriate use. J. Clin. Epidemiol. 99, 53–63.

Cooper, C., Garside, R., Varley-Campbell, J., Talens-Bou, J., Booth, A., Britten, N., 2020a. "It has no meaning to me." How do researchers understand the effectiveness of literature searches? A qualitative analysis and preliminary typology of understandings. Res. Synth. Methods 11 (5), 627–640. https://doi.org/10.1002/jrsm.1426.

Cooper, C., Snowsill, T., Worsley, C., Prowse, A., O'Mara-Eves, A., Greenwood, H., Boulton, E., Strickson, A., 2020b. Handsearching had best recall but poor efficiency when exporting to a bibliographic tool: case study. J. Clin. Epidemiol. 123, 39–48. https://doi.org/10.1016/j.jclinepi.2020.03.013.

Cooper, C., Booth, A., Husk, K., Lovell, R., Frost, J., Schauberger, U., Britten, N., Garside, R., 2022. A tailored approach: A model for literature searching in complex systematic reviews. J. Inf. Sci., 01655515221114452 https://doi.org/10.1177/01655515221114452.

Cronin, M.A., George, E., 2023. The why and how of the integrative review. Organ. Res. Methods 26 (1), 168–192. https://doi.org/10.1177/1094428120935507.
Cumpston, M., Li, T., Page, M.J., Chandler, J., Welch, V.A., Higgins, J.P., Thomas, J., 2019. Updated guidance for trusted systematic reviews: a new edition of the Cochrane handbook for systematic reviews of interventions. In: The Cochrane Database of Systematic Reviews, Vol. 10, ED000142. https://doi.org/10.1002/14651858.ED000142.
Dalton, R.D.A.N., Aguinis, H., Dalton, C.M., Bosco, F.A., Pierce, C.A., 2012. Revisiting the file drawer problem in meta-analysis: an assessment of published and nonpublished correlation matrices. Pers. Psychol. 65 (2), 221–249. https://doi.org/10.1111/j.1744-6570.2012.01243.x.
Daniels, K., 2019. Guidance on conducting and reviewing systematic reviews (and meta-analyses) in work and organizational psychology. Eur. J. Work Organ. Psy. 28 (1), 1–10. https://doi.org/10.1080/1359432X.2018.1547708.
Denyer, D., Tranfield, D., 2006. Using qualitative research synthesis to build an actionable knowledge base. Manag. Decis. 44 (2), 213–227. https://doi.org/10.1108/00251740610650201.
Denyer, D., Tranfield, D., 2009. Producing a systematic review. In: The Sage Handbook of Organizational Research Methods. Sage Publications Ltd., pp. 671–689
Dicks, L.V., Walsh, J.C., Sutherland, W.J., 2014. Organising evidence for environmental management decisions: a '4S' hierarchy. Trends Ecol. Evol. 29 (11), 607–613. https://doi.org/10.1016/j.tree.2014.09.004.
Durach, C.F., Kembro, J., Wieland, A., 2017. A new paradigm for systematic literature reviews in supply chain management. J. Supply Chain Manag. 53 (4), 67–85.
Egger, M., Smith, G.D., 1998. Bias in location and selection of studies. BMJ (Clin. Res. Ed.) 316 (7124), 61–66. https://doi.org/10.1136/bmj.316.7124.61.
Else, H., 2023. Abstracts written by ChatGPT fool scientists. Nature 613 (7944), 423. https://doi.org/10.1038/d41586-023-00056-7.
Elsevier, 2024. What are Scopus subject area categories and ASJC codes? https://service.elsevier.com/app/answers/detail/a_id/12007/supporthub/scopus/.
Frandsen, T., Gildberg, F., Tingleff, E., 2019. Searching for qualitative health research required several databases and alternative search strategies: a study of coverage in bibliographic databases. J. Clin. Epidemiol. 114. https://doi.org/10.1016/j.jclinepi.2019.06.013.
Frants, V.I., Shapiro, J., Taksa, I., Voiskunskii, V.G., 1999. Boolean search: current state and perspectives. J. Am. Soc. Inf. Sci. 50 (1), 86–95.
Gabelica, M., Bojčić, R., Puljak, L., 2022. Many researchers were not compliant with their published data sharing statement: a mixed-methods study. J. Clin. Epidemiol. 150, 33–41. https://doi.org/10.1016/j.jclinepi.2022.05.019.
Galactica, 2022. Galactica. Limitations. https://galactica.org/mission/.
Garritty, C., Gartlehner, G., Nussbaumer-Streit, B., King, V.J., Hamel, C., Kamel, C., Affengruber, L., Stevens, A., 2021. Cochrane rapid reviews methods group offers evidence-informed guidance to conduct rapid reviews. J. Clin. Epidemiol. 130, 13–22. https://doi.org/10.1016/j.jclinepi.2020.10.007.
Grant, M.J., Booth, A., 2009. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Inf. Libr. J. 26 (2), 91–108.
Greenhalgh, T., 2021. Miasmas, mental models and preventive public health: some philosophical reflections on science in the COVID-19 pandemic. Interface Focus 11 (6), 20210017. https://doi.org/10.1098/rsfs.2021.0017.
Grimaldi, G., Ehrler, B., 2023. AI et al.: machines are about to change scientific publishing forever. ACS Energy Lett. 8 (1), 878–880. https://doi.org/10.1021/acsenergylett.2c02828.
Gruner, R.L., Minunno, R., 2023. Theorizing across boundaries: how to conduct a 'breakout' literature review. Int. J. Manag. Rev. 26 (3), 331–343. https://doi.org/10.1111/ijmr.12356.
Guimarães, N.S., Ferreira, A.J.F., Silva, R. de C.R., de Paula, A.A., Lisboa, C.S., Magno, L., Ichiara, M.Y., Barreto, M.L., 2022. Deduplicating records in systematic reviews: there are free, accurate automated ways to do so. J. Clin. Epidemiol. 152, 110–115.
Gusenbauer, M., 2019. Google scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. Scientometrics 118 (1), 177–214.
Gusenbauer, M., 2021. The age of abundant scholarly information and its synthesis–A time when 'just google it' is no longer enough. Res. Synth. Methods 12 (6), 684–691.
Gusenbauer, M., 2022. Search where you will find most: comparing the disciplinary coverage of 56 bibliographic databases. Scientometrics 127 (5), 2683–2745.
Gusenbauer, M., 2023. Audit AI search tools now, before they skew research. Nature 617, 439. https://doi.org/10.1038/d41586-023-01613-w.
Gusenbauer, M., 2024a. Beyond google scholar, scopus, and web of science: an evaluation of the backward and forward citation coverage of 59 databases' citation indices. Res. Synth. Methods 15 (5), 802–817.
Gusenbauer, M., 2024b. Searchsmart.org: guiding researchers to the best databases and search systems for systematic reviews and beyond. Res. Synth. Methods. https://doi.org/10.1002/jrsm.1746.
Gusenbauer, M., Haddaway, N.R., 2020. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google scholar, PubMed, and 26 other resources. Res. Synth. Methods 11 (2), 181–217.
Gusenbauer, M., Haddaway, N.R., 2021. What every researcher should know about searching–clarified concepts, search advice, and an agenda to improve finding in academia. Res. Synth. Methods 12 (2), 136–147.
Haddaway, N.R., Westgate, M.J., 2019. Predicting the time needed for environmental systematic reviews and systematic maps. Conserv. Biol. 33 (2), 434–443. https://doi.org/10.1111/cobi.13231.
Haddaway, N.R., Grainger, M.J., Gray, C.T., 2022a. Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching. Res. Synth. Methods 13 (4), 533–545. https://doi.org/10.1002/jrsm.1563.
Haddaway, N.R., Page, M.J., Pritchard, C.C., McGuinness, L.A., 2022b. PRISMA2020: an R package and shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. Campbell Syst. Rev. 18 (2), e1230.
Harrison, J.S., Banks, G.C., Pollack, J.M., O'Boyle, E.H., Short, J., 2017. Publication bias in strategic management research. J. Manag. 43 (2), 400–425.
Harzing, A.-W., 2007. Publish or perish. https://harzing.com/resources/publish-or-perish.
Havránek, T., Stanley, T.D., Doucouliagos, H., Bom, P., Geyer-Klingeberg, J., Iwasaki, I., Reed, W.R., Rost, K., van Aert, R.C.M., 2020. Reporting guidelines for meta-analysis in economics. J. Econ. Surv. 34 (3), 469–475. https://doi.org/10.1111/joes.12363.
Heaven, Will D., 2022. Why Meta's latest large language model survived only three days online. MIT Technol. Rev. https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science/.
Hiebl, M.R.W., 2023. Sample selection in systematic literature reviews of management research. Organ. Res. Methods 26 (2), 1094428120986851. https://doi.org/10.1177/1094428120986851.
Higgins, J.P.T., López-López, J.A., Becker, B.J., Davies, S.R., Dawson, S., Grimshaw, J.M., McGuinness, L.A., Moore, T.H.M., Rehfuess, E.A., Thomas, J., Caldwell, D.M., 2019. Synthesising quantitative evidence in systematic reviews of complex health interventions. BMJ Glob. Health 4 (Suppl. 1), e000858.
Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A., 2020. Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons.
Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A., 2023. Cochrane Handbook for Systematic Reviews of Interventions Version 6.4. John Wiley & Sons. Available from. www.training.cochrane.org/handbook.
Hirt, J., Nordhausen, T., Fuerst, T., Ewald, H., group, Tarc. study, Appenzeller-Herzog, C., 2023. The TARCiS statement: guidance on terminology, application, and reporting of citation searching. MedRxiv 2023 (10), 25.23297543. https://doi.org/10.1101/2023.10.25.23297543.
Hirt, J., Nordhausen, T., Fuerst, T., Ewald, H., Appenzeller-Herzog, C., 2024. Guidance on terminology, application, and reporting of citation searching: the TARCiS statement. Bmj 385.
Hope, T., Downey, D., Etzioni, O., Weld, D.S., Horvitz, E., 2022. A computational inflection for scientific discovery. In: Proceedings of ACM Conference (Conference'17), Vol. 1, Issue 1. Association for computing machinery. http://arxiv.org/abs/2205.02007.
Hosseini, M., Rasmussen, L.M., Resnik, D.B., 2023. Using AI to write scholarly publications. In: Accountability in Research, pp. 1–9. https://doi.org/10.1080/08989621.2023.2168535.
Huisman, J., Smits, J., 2017. Duration and quality of the peer review process: the author's perspective. Scientometrics 113 (1), 633–650. https://doi.org/10.1007/s11192-017-2310-5.
Hutson, M., 2022. Could AI help you to write your next paper? Nature 611 (7934), 192–193. https://doi.org/10.1038/d41586-022-03479-w.
Ibrahim, H., Liu, F., Zaki, Y., Rahwan, T., 2024. Google Scholar is Manipulatable.
Ioannidis, J.P.A., 2016. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q. 94 (3), 485–514.
Issaiy, M., Ghanaati, H., Kolahi, S., Shakiba, M., Jalali, A.H., Zarei, D., Kazemian, S., Avanaki, M.A., Firouznia, K., 2024. Methodological insights into ChatGPT's screening performance in systematic reviews. BMC Med. Res. Methodol. 24 (1), 78.
Jones, B.F., 2009. The burden of knowledge and the "death of the renaissance man": is innovation getting harder? Rev. Econ. Stud. 76 (1), 283–317.
Jones, O., Gatrell, C., 2014. Editorial: the future of writing and reviewing for IJMR. Int. J. Manag. Rev. 16 (3), 249–264. https://doi.org/10.1111/ijmr.12038.
Khalil, H., Pollock, D., McInerney, P., Evans, C., Moraes, E.B., Godfrey, C.M., Alexander, L., Tricco, A., Peters, M.D.J., Pieper, D., 2024. Automation tools to support undertaking scoping reviews. Res. Synth. Methods. https://doi.org/10.1002/jrsm.1731.
Konno, K., Pullin, A.S., 2020. Assessing the risk of bias in choice of search sources for environmental meta-analyses. Res. Synth. Methods 11 (5), 698–713. https://doi.org/10.1002/jrsm.1433.
Kraus, S., Breier, M., Lim, W.M., Dabić, M., Kumar, S., Kanbach, D., Mukherjee, D., Corvello, V., Piñeiro-Chousa, J., Liguori, E., Palacios-Marqués, D., Schiavone, F., Ferraris, A., Fernandes, C., Ferreira, J.J., 2022. Literature reviews as independent studies: guidelines for academic practice. Rev. Manag. Sci. 16 (8), 2577–2595. https://doi.org/10.1007/s11846-022-00588-8.
Kugley, S., Wade, A., Thomas, J., Mahood, Q., Jørgensen, A.-M.K., Hammerstrøm, K., Sathe, N., 2017. Searching for studies: a guide to information retrieval for Campbell systematic reviews. Campbell Syst. Rev. 13 (1), 1–73. https://doi.org/10.4073/cmg.2016.1.
Kunisch, S., Menz, M., Bartunek, J.M., Cardinal, L.B., Denyer, D., 2018. Feature topic at organizational research methods: how to conduct rigorous and impactful literature reviews? Organ. Res. Methods 21 (3), 519–523.
Kunisch, S., Denyer, D., Bartunek, J.M., Menz, M., Cardinal, L.B., 2023. Review research as scientific inquiry. Organ. Res. Methods 26 (1), 3–45. https://doi.org/10.1177/10944281221127292.
Lakens, D., Hilgard, J., Staaks, J., 2016. On the reproducibility of meta-analyses: six practical recommendations. BMC Psychol. 4 (1), 1–10. https://doi.org/10.1186/s40359-016-0126-3.
Leclercq, V., Beaudart, C., Ajamieh, S., Rabenda, V., Tirelli, E., Bruyère, O., 2019. Meta-analyses indexed in PsycINFO had a better completeness of reporting when they mention PRISMA. J. Clin. Epidemiol. 115, 46–54. https://doi.org/10.1016/j.jclinepi.2019.06.014.

Leeflang, M.M.G., Scholten, R.J.P.M., Rutjes, A.W.S., Reitsma, J.B., Bossuyt, P.M.M., 2006. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. J. Clin. Epidemiol. 59 (3), 234–240.
Lefebvre, C., Glanville, J., Briscoe, S., Featherstone, R., Littlewood, A., Marshall, C., Metzendorf, M.-I., Noel-Storr, A., Paynter, R., Radar, T., Thomas, J., Wieland, L., 2023. Technical supplement to chapter 4: searching for and selecting studies. In: Cochrane Handbook for Systematic Reviews of Interventions Version 6.3.
Lefebvre, Carol, Glanville, J., Briscoe, S., Littlewood, A., Marshall, C., Metzendorf, M., Noel-Storr, A., Rader, T., Shokraneh, F., Thomas, J., 2019. Searching for and selecting studies. In: Cochrane Handbook for Systematic Reviews of Interventions, pp. 67–107.
Levitt, H.M., Bamberg, M., Creswell, J.W., Frost, D.M., Suárez-orozco, C., Appelbaum, M., Cooper, H., Kline, R., Mayo.Wilson, E., Nezu, A., & Rao, S., 2018. Reporting standards for qualitative research in psychology: the APA publications and communications board task force report. Am. Psychol. 1 (2), 26–46.
Linder, S.K., Kamath, G.R., Pratt, G.F., Saraykar, S.S., Volk, R.J., 2015. Citation searches are more sensitive than keyword searches to identify studies using specific measurement instruments. J. Clin. Epidemiol. 68 (4), 412–417.
Logullo, P., MacCarthy, A., Kirtley, S., Collins, G.S., 2020. Reporting guideline checklists are not quality evaluation forms: they are guidance for writing. Health Sci. Rep. 3 (2), e165. https://doi.org/10.1002/hsr2.165.
Lowe, M.S., Maxson, B.K., Stone, S.M., Miller, W., Snajdr, E., Hanna, K.A., 2017. The Boolean Is Dead, Long Live the Boolean!. In: Natural language versus boolean searching in introductory undergraduate instruction.
Maddux, W.W., Lu, J.G., Affinito, S.J., Galinsky, A.D., 2021. Multicultural experiences: a systematic review and new theoretical framework. Acad. Manag. Ann. 15 (2), 345–376. https://doi.org/10.5465/annals.2019.0138.
Martín-Martín, A., Thelwall, M., Orduna-Malea, E., Delgado López-Cózar, E., 2021. Google scholar, Microsoft academic, Scopus, dimensions, web of science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. Scientometrics 126 (1), 871–906.
McGowan, J., Sampson, M., 2005. Systematic reviews need systematic searchers. J. Med. Libr. Assoc.: JMLA 93 (1), 74–80.
McGowan, J., Sampson, M., Salzwedel, D.M., Cogo, E., Foerster, V., Lefebvre, C., 2016. PRESS peer review of electronic search strategies: 2015 guideline statement. J. Clin. Epidemiol. 75, 40–46. https://doi.org/10.1016/j.jclinepi.2016.01.021.
Meert, D., Torabi, N., Costella, J., 2016. Impact of librarians on reporting of the literature searching component of pediatric systematic reviews. J. Med. Libr. Assoc.: JMLA 104 (4), 267.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Group, and the P, 2009. Reprint—preferred reporting items for systematic reviews and Meta-analyses: the PRISMA statement. Phys. Ther. 89 (9), 873–880. https://doi.org/10.1093/ptj/89.9.873.
Moore, R.A., Fisher, E., Eccleston, C., 2022. Systematic reviews do not (yet) represent the 'gold standard' of evidence: a position paper. Eur. J. Pain 26 (3), 557–566.
Munn, Z., Stern, C., Aromataris, E., Lockwood, C., Jordan, Z., 2018. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Med. Res. Methodol. 18 (1), 1–9.
Nama, N., Hennawy, M., Barrowman, N., O'Hearn, K., Sampson, M., McNally, J.D., 2021. Successful incorporation of single reviewer assessments during systematic review screening: development and validation of sensitivity and work-saved of an algorithm that considers exclusion criteria and count. Syst. Rev. 10, 1–10.
O'Keefe, H., Rankin, J., Wallace, S.A., Beyer, F., 2023. Investigation of text-mining methodologies to aid the construction of search strategies in systematic reviews of diagnostic test accuracy-a case study. Res. Synth. Methods 14 (1), 79–98. https://doi.org/10.1002/jrsm.1593.
Paez, A., 2017. Gray literature: an important resource in systematic reviews. J. Evid. Based Med. 10 (3), 233–240. https://doi.org/10.1111/jebm.12266.
Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst. Rev. 10 (1), 1–11.
Papaioannou, D., Sutton, A., Carroll, C., Booth, A., Wong, R., 2010. Literature searching for social science systematic reviews: consideration of a range of search techniques. Health Inf. Libr. J. 27 (2), 114–122.
Paré, G., Trudel, M.-C., Jaana, M., Kitsiou, S., 2015. Synthesizing information systems knowledge: A typology of literature reviews. Inf. Manag. 52 (2), 183–199. https://doi.org/10.1016/j.im.2014.08.008.
Parisi, V., Sutton, A., 2024. The role of ChatGPT in developing systematic literature searches: an evidence summary. J. EAHIL 20 (2 SE-Feature Articles), 30–34. https://doi.org/10.32384/jeahil20623.
Park, M., Leahey, E., Funk, R.J., 2023. Papers and patents are becoming less disruptive over time. Nature 613 (7942), 138–144.
Pham, M.T., Rajić, A., Greig, J.D., Sargeant, J.M., Papadopoulos, A., Mcewen, S.A., 2014. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res. Synth. Methods 5 (4), 371–385. https://doi.org/10.1002/jrsm.1123.
Pittaway, L., Robertson, M., Munir, K., Denyer, D., Neely, A., 2004. Networking and innovation: a systematic review of the evidence. Int. J. Manag. Rev. 5 (3–4), 137–168.
Post, C., Sarala, R., Gatrell, C., Prescott, J.E., 2020. Advancing theory with review articles. J. Manag. Stud. 57 (2), 351–376. https://doi.org/10.1111/joms.12549.
Pullin, A.S., Frampton, G., Livoreil, B., Petrokofsky, G., 2018. Guidelines and Standards for Evidence Synthesis in Environmental Management. Version 5.1.
Rethlefsen, M.L., Kirtley, S., Waffenschmidt, S., Ayala, A.P., Moher, D., Page, M.J., Koffel, J.B., Blunt, H., Brigham, T., Chang, S., Clark, J., Conway, A., Couban, R., De Kock, S., Farrah, K., Fehrmann, P., Foster, M., Fowler, S.A., Glanville, J., et al., 2021. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. J. Med. Libr. Assoc. 109 (2), 174–200. https://doi.org/10.5195/jmla.2021.962.
Ringquist, E., 2013. Meta-Analysis for Public Management and Policy. John Wiley & Sons.
Ross-White, A., Godfrey, C., 2017. Is there an optimum number needed to retrieve to justify inclusion of a database in a systematic review search? Health Inf. Libr. J. 34 (3), 217–224.
Rousseau, D.M., 2024. Reviews as research: steps in developing trustworthy synthesis. Acad. Manag. Ann. 18 (2), 395–402.
Roza, M., Van den Bosch, F.A.J., Volberda, H.W., 2011. Offshoring strategy: motives, functions, locations, and governance modes of small, medium-sized and large firms. Int. Bus. Rev. 20 (3), 314–323. https://doi.org/10.1016/j.ibusrev.2011.02.002.
Ryu, J.W., Neubert, E.M., Gonzalez-Mulé, E., July 2020. Putting the team in the driver's seat: A meta-analysis on the what, why, and when of team autonomy's impact on team effectiveness. Pers. Psychol. 2021, 1–29. https://doi.org/10.1111/peps.12468.
Salvador-Oliván, J.A., Marco-Cuenca, G., Arquero-Avilés, R., 2019. Errors in search strategies used in systematic reviews and their effects on information retrieval. J. Med. Libr. Assoc.: JMLA 107 (2), 210–221. https://doi.org/10.5195/jmla.2019.567.
Sánchez-Gil, S., Gorraiz, J., Melero-Fuentes, D., 2018. Reference density trends in the major disciplines. J. Informet. 12 (1), 42–58. https://doi.org/10.1016/j.joi.2017.11.003.
Sarkis-Onofre, R., Catalá-López, F., Aromataris, E., Lockwood, C., 2021. How to properly use the PRISMA statement. Syst. Rev. 10 (1), 117. https://doi.org/10.1186/s13643-021-01671-z.
Schellinger, J., Sewell, K., Bloss, J.E., Ebron, T., Forbes, C., 2021. The effect of librarian involvement on the quality of systematic reviews in dental medicine. PloS One 16 (9), e0256833.
Sharma, G., Bansal, P., 2020. Partnering up: including managers as research partners in systematic reviews. Organ. Res. Methods 26 (2), 262–291. https://doi.org/10.1177/1094428120965706.
Simsek, Z., Fox, B., Heavey, C., 2021. Systematicity in organizational research literature reviews: A framework and assessment. Organ. Res. Methods 26 (2), 292–321. https://doi.org/10.1177/10944281211008652.
Simsek, Z., Fox, B.C., Heavey, C., Liu, S., 2024. Methodological rigor in management research reviews. J. Manag., 01492063241237222 https://doi.org/10.1177/01492063241237222.
Stanley, T.D., Doucouliagos, H., 2012. Meta-regression Analysis in Economics and Business. Routledge.
Steel, P., Beugelsdijk, S., Aguinis, H., 2021. The anatomy of an award-winning meta-analysis: recommendations for authors, reviewers, and readers of meta-analytic reviews. J. Int. Bus. Stud. 52 (1), 23–44. https://doi.org/10.1057/s41267-020-00385-z.
Stevinson, C., Lawlor, D.A., 2004. Searching multiple databases for systematic reviews: added value or diminishing returns? Complement. Ther. Med. 12 (4), 228–232. https://doi.org/10.1016/j.ctim.2004.09.003.
Susnjak, T., 2022. ChatGPT: The End of Online Exam Integrity?, pp. 1–21. http://arxiv.org/abs/2212.09292.
Terwee, C.B., Jansma, E.P., Riphagen, I.I., de Vet, H.C.W., 2009. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual. Life Res. 18 (8), 1115–1123. https://doi.org/10.1007/s11136-009-9528-5. : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation.
Thelwall, M., 2017. Avoiding obscure topics and generalising findings produces higher impact research. Scientometrics 110 (1), 307–320.
Thelwall, M., Kousha, K., Stuart, E., Makita, M., Abdoli, M., Wilson, P., Levitt, J., 2023. In which fields are citations indicators of research quality? J. Assoc. Inf. Sci. Technol. 74 (8), 941–953.
Tranfield, D., Denyer, D., Smart, P., 2003. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br. J. Manag. 14 (3), 207–222.
Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K., Colquhoun, H., Kastner, M., Levac, D., Ng, C., Sharpe, J.P., Wilson, K., Kenny, M., Warren, R., Wilson, C., Stelfox, H.T., Straus, S.E., 2016. A scoping review on the conduct and reporting of scoping reviews. BMC Med. Res. Methodol. 16 (1), 1–10. https://doi.org/10.1186/s12874-016-0116-4.
Valentine, J.C., Pigott, T.D., Rothstein, H.R., 2010. How many studies do you need? A primer on statistical power for meta-analysis. J. Educ. Behav. Stat. 35 (2), 215–247. https://doi.org/10.3102/1076998609346961.
Walpole, S.C., 2019. Including papers in languages other than English in systematic reviews: important, feasible, yet often omitted. J. Clin. Epidemiol. 111, 127–134. https://doi.org/10.1016/j.jclinepi.2019.03.004.
Waltman, L., Noyons, E., 2018. Bibliometrics for Research Management and Research Evaluation: A Brief Introduction (CWTS Meaningful metrics).
Wang, S., Scells, H., Koopman, B., Zuccon, G., 2023. Can ChatGPT Write a Good Boolean Query for Systematic Review Literature Search?. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval https://api.semanticscholar.org/CorpusID:256627169.
Wanyama, S.B., McQuaid, R.W., Kittler, M., 2022. Where you search determines what you find: the effects of bibliographic databases on systematic reviews. Int. J. Soc. Res. Methodol. 25 (3), 409–422.

Wilczynski, N.L., Haynes, R.B., Eady, A., Haynes, B., Marks, S., McKibbon, A., Morgan, D., Walker-Dilks, C., Walter, S., Werre, S., Wilczynski, N., Wong, S., 2004. Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. BMC Med. 2, 1–5. https://doi.org/10.1186/1741-7015-2-23.

Wood, J. "Andy", 2008. Methodology for dealing with duplicate study effects in a meta-analysis. Organ. Res. Methods 11 (1), 79–95. https://doi.org/10.1177/1094428106296638.

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), 1–8. https://doi.org/10.1186/1471-2288-14-73.

Xiao, Y., Watson, M., 2019. Guidance on conducting a systematic literature review. J. Plan. Educ. Res. 39 (1), 93–112. https://doi.org/10.1177/0739456X17723971.

Yeung, S.K., Yay, T., Feldman, G., 2021. Action and inaction in moral judgments and decisions: meta-analysis of omission bias omission-commission asymmetries. Personal. Soc. Psychol. Bull.

Zhou, Z.Q., Tse, T.H., Witheridge, M., 2021. Metamorphic robustness testing: exposing hidden defects in citation statistics and journal impact factors. IEEE Trans. Softw. Eng. 47 (6), 1164–1183. https://doi.org/10.1109/TSE.2019.2915065.

Zupic, I., Čater, T., 2015. Bibliometric methods in management and organization. Organ. Res. Methods 18 (3), 429–472.

Michael Gusenbauer is an Assistant Professor at the Institute of Innovation Management at Johannes Kepler University Linz, Austria. He received his Ph.D. from Johannes Kepler University Linz, a diploma degree from WU Vienna, and has previously held post-doctoral positions at the Technical University of Munich and the University of Innsbruck. Michael is the founder of *Search Smart*, a website to compare academic databases. His research has been published in journals such as *Nature*, *Journal of Management Studies*, *Family Business Review*, *Research Synthesis Methods*, and *Scientometrics*.

Sebastian Peter Gauster is a researcher and lecturer at the Innsbruck University School of Management. Sebastian holds a Ph.D. in management and a diploma degree in international business and economics from the University of Innsbruck. In his research, he puts a focus on agility, agile transformations, and innovation management. Further, he is interested in systematic literature reviews, meta-analysis, and bibliometric studies. He has published in *Management Decision* and has regularly presented at conferences such as the *Academy of Management Conference* and the *European Academy of Management Conference*.