

HOW TO

How to... conduct a systematic review to inform practice, policy and research in healthcare simulation

Sinéad Lydon^{1,2} and Caoimhe Madden^{2,3}¹Department of General Practice, School of Medicine, University of Galway, Galway, Ireland²Irish Centre for Applied Patient Safety and Simulation (ICAPSS), School of Medicine, University of Galway, Galway, Ireland³Discipline of Medical Education, School of Medicine, University of Galway, Galway, Ireland**Corresponding author:** Sinéad Lydon, sinead.lydon@universityofgalway.ie<https://johs.org.uk/article/doi/10.54531/JGPF2091>

ABSTRACT

The importance of, and resource requirements for, healthcare simulation (HS) demand that best practice in its design, delivery and assessment are established and adhered to in order to ensure optimal outcomes for learners, systems and the patients they serve. Establishing what constitutes best practice in the implementation of HS is therefore crucial. Systematic reviews (SRs) are a form of evidence synthesis that use rigorous methods to identify and combine all relevant literature on a research topic or question to offer reliable, considered guidance that informs understanding, day-to-day practice, future research and policy. Conducting a SR can be a challenging and resource-intensive endeavour. This article offers a step-by-step guide, contextualised for the field of HS, which will support researchers throughout all stages of the SR process, from articulating an appropriate review question, through identifying and selecting relevant research studies, to preparing a manuscript and overcoming common challenges.

Key points

- Given the importance, and resource requirements, of Healthcare Simulation, establishing and adhering to evidence-based practice is crucial but challenging within this evolving, multidisciplinary field.
- A variety of evidence synthesis methodologies exist, varying in focus, paradigm and quality, to support the bringing together of diverse bodies of literature to generate understanding, inform theory or evidence-based guidance, and/or identify knowledge gaps in a research literature.
- Systematic reviews (SRs) are one form of evidence synthesis that employ explicit, reproducible methods to identify relevant primary research studies, critically appraise their quality and combine their results to provide meaningful and reliable guidance to support understanding, day-to-day practice, future research and/or policy.
- SRs can provide guidance on various issues such as clarifying the impact of an intervention, individuals' experiences of an event or intervention, how constructs should be measured, resource usage or costs, current expert opinion or policy guidance or establishing how research should be designed or conducted.

Submission Date: 02 November 2025

Accepted Date: 22 May 2026

Published Date: 22 June 2026

- Researchers conducting a SR will establish the importance of conducting the review and consider the necessary resources, articulate an appropriate review question, prepare a protocol, identify relevant literature, screen and select studies for inclusion, extract data from studies, critically appraise studies, synthesise the data across studies and prepare a written manuscript.
- Delivering an impactful SR requires considerable time, manpower and expertise along with an awareness of, and adherence to, established best practice.

Evidence synthesis refers to a process of collating information from multiple sources (e.g. peer-reviewed articles, government reports) to summarise a body of evidence [1], providing a more comprehensive understanding of a particular topic or answer to a research question, than a single study could offer [2]. The various evidence synthesis methodologies are diverse, differing in terms of focus, paradigm and quality (see Table 1). Most are characterised by a rigorous, systematic and transparent approach to generating research evidence and understanding [1], with the search and synthesis processes explicitly documented to allow replication, facilitate updates and assist in identifying potential bias [3]. Evidence synthesis is often a crucial step in the use of research for personal and public decision-making [1], applied to support understanding of a phenomenon or lived experiences, theory development, generation of reliable recommendations informed by the best available evidence, and/or identifying knowledge gaps for future research [4,5].

Evidence synthesis methodologies are applied across diverse fields of research [16], and disciplinary variation can exist in terms of the methodologies typically applied. Traditionally, SRs have been regarded as the ‘pillar’ [17] of evidence-based medicine and widely used to inform the delivery of optimal healthcare interventions (e.g. by establishing their benefits and harms). SRs employ explicit, reproducible methods to identify relevant primary research studies, critically appraise their quality and combine their results to provide reliable answers to pre-defined research questions [5,18]. Over time, SR methods and corresponding best practice guidance have evolved to address other topics of importance to educators, clinicians, patients and policymakers beyond intervention effectiveness [19] (e.g. quality and characteristics of measurement instruments or how individuals experience an event/phenomenon and factors that influence experiences). This reflects the reality that providing high-quality, evidence-based care requires more than the delivery of effective interventions alone and that understanding the experience of an event, intervention or process can be critically important [20].

While SRs remain crucially important for informing day-to-day practice (e.g. delivery of optimal healthcare interventions), policy (e.g. development of clinical guidelines) and furthering research in healthcare, and are the focus of the current guide, it is important for researchers to be aware that the use of other evidence synthesis methodologies, which may be more fluid, theory-focused or integrative, has grown and may be more appropriate for establishing what research exists in a specific area (e.g. scoping review,

mapping review), collating evidence from qualitative research studies (e.g. qualitative SR) or statistically examining the effectiveness of interventions (e.g. meta-review; see Table 1 for additional examples). Suitability of individual evidence synthesis methodologies will depend on a range of factors, including the nature of the research question, the evidence base, available resources and timeline for completion. Therefore, while we assert that the conduct of high-quality SRs is essential to supporting the development of the field of healthcare simulation (HS), and provide guidance to support researchers in their conduct, we also wish, from the outset of this guide, to support an awareness of the importance and place for, other evidence synthesis methodologies in HS, though providing granular detail on their conduct is outside the scope of the current article.

Relevance of systematic reviews to healthcare simulation

HS is widely seen as an effective method for education and training (e.g. use of simulators for developing essential technical and non-technical skills in a safe learning environment) in the health professions [21]. HS also serves important non-pedagogical functions through supporting improvement by identifying system-wide challenges and supporting redesign, coined ‘transformative simulation’ [22] (e.g. testing quality improvement initiatives, identifying latent safety threats). Although the use of simulation has been described for hundreds of years [21], the implementation of HS has grown exponentially in the 21st century, leading to a considerable increase in HS research articles and their spread across academic journals, in recent decades [6]. Recent years have therefore resulted in a critical mass of published work that enables and necessitates, meaningful evidence synthesis.

Evidence synthesis in this field is crucial for several reasons. First, given the evolving nature of HS, it can be challenging for practitioners to stay abreast of rapidly advancing developments (e.g. the increased application of transformative simulation), though engagement with HS research is crucial for supporting their work. Second, implementing HS is expensive and resource-intensive [23] but crucially important for learner [24], system and patient outcomes [25]; therefore, providing guidance for HS researchers will be useful for avoiding resource wastage and ensuring their work is grounded in evidence. Conserving resources in day-to-day HS practice, where possible, may allow the time, space and people required for the experimentation and adaptation that is crucial to the growth and full use of HS. Third, responding to frequently

Table 1: Alternative evidence synthesis methodologies that healthcare simulation researchers may consider

| | | |
|---|---|---|
| <p>Bibliometric review Focuses on quantitatively summarising trends around the characteristics of research publications on a particular topic or in a particular field. Characteristics of interest may relate to the authors (e.g. gender, country of employment), the work conducted (e.g. study design, research focus) or the research publication (e.g. citations, funding). Supports an understanding of how, where and by whom, research in the field is being produced or its use and impact.</p> <p>Healthcare simulation example: Walsh et al. [6] <i>The 100 most cited articles on healthcare simulation: a bibliometric review.</i></p> | <p>Mapping review Focuses on establishing what research exists in a specific area, categorising studies into clearly defined areas, and plotting gaps in the literature. The intent is to deliver a useful, and often visual, 'map' of a research topic to facilitate targeting of subsequent research or review questions and activity. Offers a broad overview of what has been studied in a research area and analysis of studies is typically limited and simple, focusing on grouping studies by what they have done or explored rather than appraising or interpreting these.</p> <p>Healthcare simulation example: Essex et al. [7] <i>A systematic mapping literature review of ethics in healthcare simulation and its methodological feasibility.</i></p> | <p>Meta-review (Umbrella review/overview of reviews) A systematic review (SR) of review articles. Uses a rigorous review methodology to distil the learning from existing evidence syntheses into one accessible document. Addresses the challenge of making evidence-based decisions where multiple relevant reviews exist.</p> <p>Healthcare simulation example: Decker et al. [8] <i>The impact of simulation debriefing process on learning outcomes: an umbrella review.</i></p> |
| <p>Narrative review An overview, or 'essay'-type, consideration of a particular topic, which does not typically seek to be comprehensive in its coverage of relevant literature. With an established lack of best practice guidance, narrative reviews are heavily directed and shaped by authors, including the decisions on included studies and focus of the write-up, resulting in a risk of bias and potentially invalid inferences and conclusions. Narrative reviews are often produced by influential figures or leaders in a research field.</p> <p>Healthcare simulation example: Slavinska et al. [9] <i>Narrative review of legal aspects in the integration of simulation-based education into medical and healthcare curricula.</i></p> | <p>Qualitative evidence synthesis (Qualitative systematic review; Qualitative meta-synthesis) Collates evidence from qualitative research studies, focusing on identifying themes or patterns that emerge across relevant studies with the intent of deepening the understanding of a phenomenon. Typically employs rigorous systematic review methodologies to achieve this aim.</p> <p>Healthcare simulation example: Denniston et al. [10] <i>Learning outcomes for communication skills across the health professions: a systematic literature review and qualitative synthesis.</i></p> | <p>Rapid review A systematic review of a body of research where the work is completed within a short timescale. Often employed in instances where an urgent need to support evidence-based decision making is identified. Time constraints are typically navigated by abbreviating systematic review steps such as reducing the extent of the search processes (e.g. fewer electronic databases), conducting only 'essential' data extraction, employing narrow eligibility criteria and/or limiting the study appraisal. Reducing the rigour may threaten the validity of conclusions and assertions. However, the provision of speedy syntheses of important bodies of evidence may be crucial in certain instances.</p> <p>Healthcare simulation example: Seabrooke and Seabrooke [11] <i>In situ clinical education of frontline healthcare providers in under-resourced areas: a rapid review.</i></p> |
| <p>Scoping review Narratively evaluates the nature and extent of a body of literature, allowing for gaps and future direction to be identified. Scoping reviews typically share a rigorous methodology with systematic review, but their focus and aim is quite different. Instead of answering a specific, bounded review question, scoping reviews typically provide an understanding of how much research has been conducted in an area and what it has considered. As the focus is not typically on the findings of studies, the quality of included studies is not usually appraised and these reviews are not typically conducted to, or suitable for, informing decision making or practice though there is typically some consideration of the findings and outcomes reported within studies.</p> <p>Healthcare simulation example: Smallheer et al. [12] <i>A scoping review of the priority of diversity, inclusion and equity in health care simulation.</i></p> | <p>Systematic review A systematic review uses standardised methodologies to methodically identify, appraise and combine all studies on a particular research topic. Systematic reviews have been widely applied in healthcare for decades, and have supported the focus on delivering evidence-based care. While earlier systematic reviews tended to focus on the effectiveness of treatments, over time, the application of systematic reviews within healthcare has broadened. Systematic reviews are now often used to bring together existing data on topics other than intervention effectiveness including individuals' experiences, measurement and cost or resource usage.</p> <p>Healthcare simulation example: Mahdi et al. [13] <i>systematic review on the current state of disaster preparation Simulation Exercises (SimEx).</i></p> | <p>Systematic review with meta-analysis A systematic review with statistical synthesis of data from included studies which offers a precise estimation of an effect or outcome observed in the research. The focus is typically on a narrow, homogenous group of studies and usually the degree of effectiveness of an intervention. Meta-analysis is challenging, and indeed often inappropriate, when a body of literature is heterogeneous and study designs or measurement vary.</p> <p>Healthcare simulation example: Mazzone et al. [14] <i>A systematic review and meta-analysis on the impact of proficiency-based progression simulation training on performance outcomes.</i></p> |

Note: The descriptions within this table have been informed by the work of Grant and Booth [15].

changing practice and clinical environments (e.g. COVID-19 pandemic, AI technologies) requires engagement with diverse technologies or the rapid application of HS to novel clinical scenarios, skills or systems and HS research can support effective responses in such instances [26]. Providing clear, accessible summaries of HS research evidence as the field matures and novel applications of simulation emerge may therefore offer a means of supporting HS educators in their work and ensuring HS is deployed to optimal effect. However, evidence synthesis may be particularly complex in this research area given its diverse disciplinary roots [27,28].

Description of the systematic review method

This article provides a step-by-step guide to conducting a SR, congruent with best practice and substantial practical experience shared by the authors, which supports novice HS researchers to understand the value, process and resource requirements of SRs, overcome common challenges and conduct their own SR to advance HS theory, practice, policy and research. An overview of the steps involved in conducting a SR is presented in Figure 1.

Step 1: establish the necessity and type of review to be conducted and consider resources required

Before beginning the practical work of a SR, a researcher should consider four crucial questions. First, is a review of the literature on this topic necessary and will it be of value? To address this, researcher(s) must decide if the question being asked is one of importance currently [29], will valuably guide future research, and whether there is a need for evidence to support decision-making and practice. Further, researchers must determine whether the findings will be novel [29]. This can be achieved by conducting simple literature searches (e.g. via Google Scholar) to identify any relevant published reviews, then considering their methodological quality, completeness and recency. Where a good-quality SR has been conducted in the previous 5–8 years, it is unlikely that a new review will add significant value except where there have been substantive developments or research activity.

Second, is a SR the right review methodology for achieving the aims and addressing the research

question(s)? Confusion over what constitutes a SR, and a mischaracterisation of reviews as such, is pervasive [20]. Other forms of evidence synthesis (see Table 1) have emerged and grown in popularity [19], and researchers should determine whether an alternative review methodology might be more suitable. SR is considered appropriate when focused on: collating international evidence, establishing current practice or differences in practice, determining priorities for future research, exploring conflicting results and delivering clear guidance to support decision-making [20]. The *Right Review* tool [30] may be helpful for offering direction on appropriate review design.

Third, what type of SR is appropriate in this instance? Types of SRs likely to be relevant to HS researchers include:

- *Effectiveness reviews*: assessing the benefits and/or harms of an intervention or practice [19,20];
- *Experiential or qualitative reviews*: improving understanding of interventions, their implementation or impact and/or exploring stakeholder interpretation of an intervention/event/experience [19,20];
- *Costs/economic evaluation reviews*: assessing the costs and resource usage of an intervention/service/experience/process/procedure [20];
- *Measurement/psychometric reviews*: exploring the characteristics of measurement instruments and clarifying best practice in assessing a construct [19,20];
- *Expert opinion/policy reviews*: synthesising documents describing guidelines, policies or protocols, sometimes alongside relevant original research, to clarify best practice [20]; and
- *Methodology reviews*: considering aspects of study design or conduct with a focus on research methods and/or quality [20].

Table 2 introduces two example SRs (an *effectiveness review* and a *measurement review*) of HS research. These reviews have not been conducted but are offered as examples to support readers in understanding and implementing the guidance offered herein.

Fourth, does the researcher(s) have access to the resources (e.g. time, people and expertise) available to enable completion of a methodologically rigorous review?

Figure 1: Steps involved in conducting a systematic review

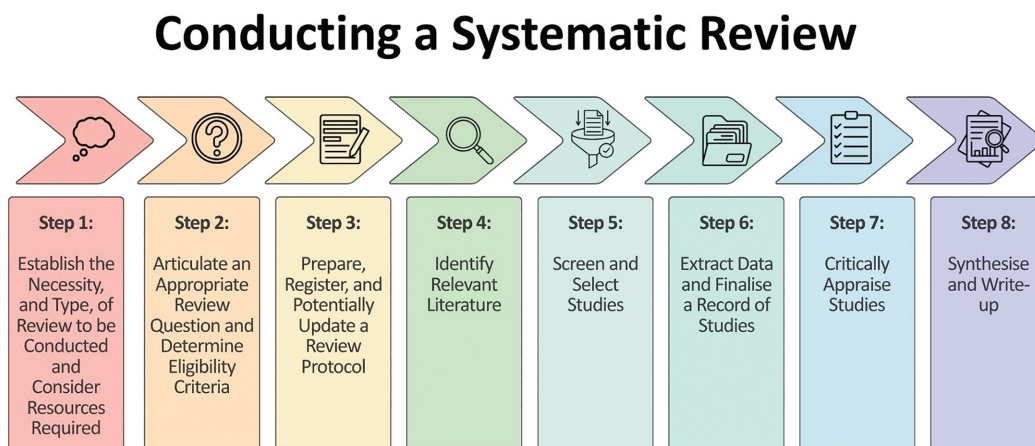


Table 2: Summary of key guidance, and example healthcare simulation systematic reviews illustrating each step of the SR process

| | Key recommendations | Example review 1 (effectiveness review) | Example review 2 (measurement review) |
|--|--|---|--|
| Articulate an appropriate review question | <ul style="list-style-type: none"> Apply appropriate framework to guide development of a clear, targeted review question. If posing multiple review questions, clarify the hierarchy or prioritisation of these. Apply FINER criteria [32] to confirm review question appropriateness. Once review question is final, develop detailed eligibility criteria. Establish appropriate scope of SR, with consideration or involvement of intended end-users and their needs. | <p><i>Research question:</i> what is the efficacy of using simulation to teach paediatric lumbar puncture among paediatric trainees?</p> <p><i>PICO framework for intervention reviews [33]:</i> Population: paediatric trainees Intervention: simulation-based training targeting performance of paediatric lumbar puncture Comparator: training as usual or no training Outcome: performance accuracy, successful/traumatic lumbar puncture in clinical setting</p> <p><i>Inclusion criteria:</i> Studies that: are published in a peer-reviewed journal; Report original research; Employ the use of simulation education for training LP procedure*; Focus on training paediatric trainees; Report outcome data on the impact of the training with the use of at least one formal measurement tool. *These criteria could be further limited to specific simulation modalities</p> <p><i>Exclusion criteria:</i> Non-empirical studies (e.g. editorials, commentaries); Studies using simulation for undergraduate education; Studies that recruit a wider population of HCPs, where it is not possible to extract the data for paediatric trainees in isolation; Studies that do not report outcome data.</p> | <p><i>Research question:</i> how has the impact of in-situ simulation to identify latent safety threats been measured in the research literature?</p> <p><i>PICO framework for measurement/ psychometric reviews [34,35]:</i> Population: healthcare professionals participating in hospital-based high-fidelity in-situ simulation(s) focused on identifying latent safety threats Instrument: quantitative and/or qualitative measures applied Construct: impact of in-situ simulation focused on identifying latent safety threats Outcomes: characteristics of measures (e.g. number of items, validity, reliability, usability)</p> <p><i>Inclusion criteria:</i> Studies that: are published in a peer-reviewed journal; Report original research; Assess the impact of at least one session of high-fidelity simulation in a healthcare setting implemented with the intention of identifying latent safety threats using at least one explicit measurement tool; Provide detail on at least psychometric characteristic of the applied measure (e.g. reliability, validity).</p> <p><i>Exclusion criteria:</i> Non-empirical studies (e.g. editorials, commentaries); Studies conducted in a simulation laboratory setting; Studies using in-situ simulation without a reported focus on identifying latent safety threats; Studies that do not implement a formal measurement tool; Studies that do not provide detail on any psychometric characteristics of the measure used.</p> |
| Develop a review protocol | <ul style="list-style-type: none"> Assemble an appropriate review team. Develop detailed protocol in adherence with the PRISMA-P guidelines [36]. Pilot the work of key steps to confirm clarity, feasibility and appropriateness of plans. Prospectively register the final SR protocol online within an appropriate repository. Update protocol if a clear need arises and detail changes made within SR manuscript. | | |
| Identify relevant literature | <ul style="list-style-type: none"> Develop overall search strategy and appropriate search terms for individual electronic databases, supported by engagement with relevant literature, piloting of searches and the advice of a research librarian. Apply the PRESS checklist [37] to evaluate plans. Adhere to Atkinson [38] and Colleagues' reporting standards for literature searches. Search three to five electronic databases. Consider the use of supplemental search tactics (particularly reference list screening) based on resources available and anticipated body of research. Update searches if there is a significant interlude between searches and manuscript submission. | <p>Search 5 electronic databases (MEDLINE, CINAHL, ERIC, PsycInfo, Web of Science) and supplement with reference list screening and forwards citation chasing (based on anticipated reasonable size and spread of relevant literature)</p> <p>Search terms to be organised within three sets:</p> <p>Terms focused on Healthcare Simulation (e.g. subject headings such as 'Patient Simulation' OR 'Simulation Training' combined with free-text entries such as 'Simulat*' OR 'task adj1 trainer*1')</p> <p>Terms focused on Paediatrics (e.g. subject headings such as 'Pediatrics' combined with free-text entries such as 'p?ediatric*1' OR 'neonat*')</p> <p>Terms focused on Lumbar Puncture (e.g. subject headings such as 'Spinal Puncture' combined with free-text entries such as 'lumbar adj1 puncture*1' OR 'spinal adj1 tap*1')</p> | <p>Search 3 electronic databases (MEDLINE, CINAHL, Web of Science; based on anticipated high volume of relevant studies and specific focus on in-situ simulation).</p> <p>Search terms to be organised within two sets:</p> <p>Terms focused on In-situ simulation (e.g. subject headings such as 'Patient Simulation' OR 'High Fidelity Simulation Training' combined with free-text entries such as 'high adj1 fidelity' OR 'in adj1 situ adj1 simulation*1')</p> <p>Terms focused on identifying latent safety threats (e.g. Subject headings such as 'Patient Safety' combined with free text entries such as 'latent adj1 (safety OR error* OR hazard*')</p> <p>Terms focused on assessment of outcomes or impact (e.g. subject headings such as 'Outcome Assessment, Health Care' combined with free-text entries such as 'outcome*1' OR 'measure*')</p> |

Table 2. Continued

| | Key recommendations | Example review 1 (effectiveness review) | Example review 2 (measurement review) |
|--|--|--|---|
| Screen and select studies | <ul style="list-style-type: none"> • Use electronic reference management system to identify and remove duplicate citations. • Conduct initial screening (i.e. title and abstract screening), ideally with involvement of, and discussion between, two researchers. • Conduct second-stage screening (i.e. full-text screening), with involvement of, and discussion between, two or more researchers as necessary. • Agree final list of studies for inclusion from electronic databases searches and implement any supplemental search tactics at this stage. • Maintain careful documentation of the work at each stage of the process (e.g. reasons for exclusion of studies during full-text review). | | |
| Extract data and finalise a record of studies | <ul style="list-style-type: none"> • Select appropriate data extraction tool. • Develop appropriate data extraction form and accompanying codebook/guide. • Devote time for training researchers in use of the data extraction form. • Pilot the form on a sample of included studies with revision as necessary. • Two researchers extract data independently from each study. • Researchers compare files, resolve discrepancies, and agree a final record of data extraction. • Code extracted data via content analysis as necessary to facilitate synthesis. | <p>Data extraction items to include:</p> <ul style="list-style-type: none"> • Author, year • Country the study was conducted in • Study design • Setting • Number and level of participants • Intervention characteristics (e.g. learning outcomes, simulator type, additional learning opportunities and resources, facilitators involved, duration and timing) • Outcome measures employed (e.g. confidence/knowledge questionnaire, accuracy and timing of skill performance) • Impact of training recorded | <p>Data extraction items to include:</p> <ul style="list-style-type: none"> • Author, year • Country the study was conducted in • Study design and methodology • Setting • Number, level and profession of participants • Measure characteristics (e.g. number of items, validity, reliability, usability any identified barriers and facilitators to its' use) • Findings |
| Critically appraise included studies | <ul style="list-style-type: none"> • Devote sufficient time to identifying a critical appraisal tool appropriate for application with the types of studies that will be included that can be effectively implemented by researchers. • Allow time for training and familiarisation with the selected tool. • Ensure novice reviewers are appropriately supported in applying the tool. • Involve two researchers in completing the critical appraisal, working independently or in tandem, with discussion to agree final judgements. | <p>A relevant tool should have the capacity to consider non-randomised interventional research, and/or research evaluating educational interventions. Researchers might consider applying tools such as:</p> <ul style="list-style-type: none"> • MERSQI [39] • the Risk-of-Bias in Non-randomized Studies of Interventions tool [40] | <p>A relevant tool should have the capacity to consider non-randomised interventional research and might usefully support consideration of the psychometric characteristics of measurement tools. Researchers might consider applying tools such as:</p> <ul style="list-style-type: none"> • QuADs tool [41] • The COSMIN Risk-of-Bias tool to assess the quality of studies on measurement properties [34] |
| Synthesis and write-up | <ul style="list-style-type: none"> • Consider and select appropriate synthesis approach. • Complete synthesis of data to bring together results of included studies, supplementing textual descriptions with visual tools in the form of tables and figures, use of counting or statistics as helpful in summarising the data, subgroup analysis, thematic or content analysis or development of conceptual models as useful to 'tell the story' of the data. • Identify target journal, noting available guidance on word count presentation etc., and examine SRs previously published within it. • Write up manuscript according to PRISMA guidelines [42]. • Make use of appendices or supplementary material as possible to be comprehensive in presentation of the SR's processes and data. | | |

*Abbreviations: HCPs: Healthcare Professionals; MERSQI: Medical Education Research Study Quality Instrument; PRESS: Peer Review of Electronic Search Strategies; PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses; PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols; QuADs: Quality Assessment with Diverse Studies.

Research [31] suggests completing and publishing a review requires, on average, 67.3 weeks (not including the time required to assemble the research team and develop the review protocol) and involves a mean of five researchers. At a minimum, a SR team should include two members to enable adherence to best practice throughout (e.g. duplicate data extraction, critical appraisal), and the team must possess the requisite subject matter knowledge to allow completion

of the review to a high standard. Further, relevant experts should be consulted for specific SR processes should they fall outside the remit of expertise of the core group; for example, having a research librarian review the search strategy, a statistician advise on data analysis plan, etc. It is crucial that consideration of resources occurs at an early stage of the process, as this will avoid the non-completion of reviews or the poor conduct of important HS reviews.

Step 2: articulate an appropriate review question and determine eligibility criteria

As in any research project, turning the identified problem into a targeted research question is fundamental to a SR's success [43]. A clearly defined review question promotes clarity of thought, requiring the researcher to explicate the SR's intended purpose [44]. Further, it supports confirmation of review methodology, given that different approaches are suited to different types of questions (see Table 1) and guides the conduct of all subsequent steps of the SR (e.g. development of eligibility criteria, determination of data extraction items [44]).

A good SR question is appropriate, meaningful and capable of being posed as a single sentence addressing a primary research question, although a well-defined hierarchy of review questions is also acceptable [43]. Given the diverse nature of research foci and methodologies across healthcare fields, a great number of frameworks exist to support HS researchers in formulating research questions, with suitability dependent on the type of SR being conducted and nature of the research literature being examined. These frameworks can support the consideration of diverse questions, beyond intervention effectiveness to understanding how the application of simulation for pedagogical or non-pedagogical purposes is experienced or elucidating the nuances of how, when and for whom/where the outcomes of its application differ. Some examples include:

- *Effectiveness SRs: PICO* [33] (Population, Intervention, Comparison, Outcome) is the most used mnemonic, sometimes modified to 'PICOS' to include study design [45] or 'PICOT' to specify a relevant timeframe [46];
- *Experiential/qualitative SRs: SPIDER* [47] (Sample, Phenomenon of Interest, Design, Evaluation, Research type);
- *Costs/economic evaluation SRs: PICOC* [48] (Population, Intervention, Comparator/s, Outcomes, Context);
- *Measurement/psychometric SRs: PICO* [34,35] (Population, Instrument, Construct, Outcomes);
- *Expert opinion/policy SRs: ECLIPSE* [49] (Expectation, Client group, Location, Impact, Professionals, Service); and
- *Methodology SRs: SDMO* [4] (Study type, Data type, Method type, Outcome).

For example, in our *effectiveness* SR, the researcher(s) should consider and define the four PICO components: the Participant(s) (i.e. Paediatric trainees), the Intervention(s) (i.e. simulation-based education; this could be further limited, for example, to specific simulation modalities or by purpose of application, depending on the focus of the review), the Comparison(s) (i.e. teaching as usual or no teaching) and the Outcome(s) (i.e. performance accuracy) [50]. Table 2 also illustrates the application of PICO within an example measurement/psychometric review. The use of such frameworks assists in ensuring an appropriately targeted review question and is associated with more complex search strategies and precise search results [51]. When a review question has been developed, the FINER

criteria [32] may be applied to confirm it is: *Feasible* (i.e. of manageable scope); *Interesting* (i.e. will yield an answer important for research and practice); *Novel* (i.e. addresses a knowledge gap); *Ethical* (i.e. requires acceptable engagement with human subjects – typically not applicable to SRs) and *Relevant* (i.e. to knowledge, policy or future research).

Researchers must also consider and establish the scope of the SR (i.e. how broad or narrow its focus should be [52]) when defining their review question. Broader reviews synthesise a greater volume of evidence [18], allowing for generalisability and consistency of findings across a wider range of settings and populations [52]. However, if the proffered question is too broad (e.g. 'Is simulation beneficial?'), then the researcher may struggle to arrive at a manageable number of studies for inclusion and generating actionable recommendations will be challenging [50]. Reviews addressing narrower questions may be more feasible [18], and the increased homogeneity of included studies may support offering focused recommendations [52]. However, the question must not be too tightly bounded or there may be insufficient relevant literature to inform useful guidance [50] (e.g. 'What is the efficacy of using simulation to teach paediatric lumbar puncture to advanced nurse practitioners in Ireland?'). When establishing the optimal scope for their SR, researchers should consider the intended end-users' needs and available resources [52] and engage relevant stakeholders in such discussions.

When a review question is final, researchers must use it to establish the eligibility criteria for including/excluding studies in the SR. Inclusion criteria are characteristics (e.g. demographic, clinical, methodological and/or geographic) that studies must possess to be included [53], while exclusion criteria describe studies that are related to some degree but will not contribute useful research data and are thereby ineligible for inclusion. These criteria should flow directly from the review question(s) and be specified a priori [54]. For each PICO/mnemonic component, the researcher(s) should determine what relevant information they wish to focus on. For instance, for 'Population', what participants are they interested in? For 'Intervention', is there a modality, purpose or type of HS intervention that should be specified? For 'Outcome', what type of data or measurements will be of use? Table 2 outlines relevant eligibility criteria for the example SRs. These criteria are crucial for guiding the development of search terms and decision-making during study selection.

Step 3: prepare, register and potentially update a review protocol

Developing a protocol at the outset of the SR process ensures that all researchers have a shared understanding of plans, that they can be feasibly implemented and that the work constitutes good practice. When preparing a protocol, researchers should comply with the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) guidelines [36], which requires:

- *Rationale*: research-based justification of the need for the review and its evidence;

- *Objectives*: the review question(s) that have been developed;
- *Eligibility criteria*: guidelines on the characteristics of the studies that will be included, with decision-making justified as appropriate;
- *Information sources*: each search tactic (including where, when and how) that will be used to retrieve relevant articles;
- *Data management*: how records and data will be handled throughout including any software to be used;
- *Selection process*: how, and how many, researchers will work to identify, consider, and select studies for inclusion;
- *Data extraction*: how, and what, data will be taken from studies and by who;
- *Critical appraisal*: how methodological rigour of studies will be assessed, and resultant data will be used; and
- *Synthesis*: plans for bringing the data together to create a useful summary and to answer the review question.

Protocol development should be viewed as a ‘trailing’ process, allowing researchers to pilot plans for each step and clarify any ambiguity identified. For instance, the planned search strategy should be trialled within electronic databases to ensure the resulting returns are manageable and relevant. When researchers are satisfied that the plans are clear, relevant and achievable, a final protocol should be agreed.

The protocol should be made accessible online before formal searches have been initiated [55]. Research suggests that SRs with prospectively registered protocols have greater methodological rigour [56] and are more likely to be published in higher impact factor journals [57]. Indeed, the PRISMA guidelines require a priori protocol registration [42]. Perceived barriers may include costs, concerns that others may ‘take’ the idea or excessive workload [58]. However, SR protocols can be made available online with minimal effort and often at no cost [59], via:

- The International Prospective Register of Systematic Reviews (PROSPERO [60]);
- Research Registry [61];
- INPLASY [62];
- OSF Registries [63]; or
- Protocols.io [64].

Registration requires little work beyond that inherent in developing the protocol [55] and may save time by ensuring a clear plan is in place, thereby reducing the opportunity for errors and miscommunications [58]. Instances of ‘stealing’ of SR plans are low [65], and protocol registration may be considered a means of ‘staking a claim’ on a topic [58]. Additionally, researchers may consider publishing their protocol in a relevant journal as an article for peer review, ideally before data collection has commenced. A number of HS-specific journals (e.g. *International Journal of Healthcare Simulation*) accept review protocols as an article type. Publication of protocols may yield great exposure and engagement, as well as furthering transparency, reproducibility and accountability.

Devoting sufficient time and effort to the initial protocol will help ensure the review can be completed as planned. However, in some instances, a need to revise the protocol may arise. For example, data on a particular outcome may not be presented within included studies and a protocol may need updating to omit this data [55]. Many review registries permit records to be updated, allowing researchers to clearly document and justify changes made [59]. Any amendments to the review protocol should be carefully considered and described in the review manuscript [42].

Step 4: identify relevant literature

Once a protocol is agreed, the process of identifying studies for inclusion can begin. Typically, a SR search strategy will comprise electronic database and reference list screening but may be supplemented with additional search tactics. When composing a search strategy, researchers should carefully consider available resources, the nature of the review, where eligible studies are likely to be published and the time and effort required to enact any search tactic as compared to the likelihood of it yielding additional relevant studies. Researchers should consult Atkinson and colleagues’ [38] reporting standards for literature searches to ensure search quality, transparency and replicability.

Electronic database searches constitute the core work of searching within a SR. Databases available to researchers will vary by institution, and their appropriateness will vary depending on the review’s focus. We recommend including three to five electronic databases, in line with research suggesting most relevant articles are found within a limited number of databases [66]. Where smaller numbers of studies are expected, searching more databases may be reasonable [66]. The following databases are likely to be useful within SRs of HS [66,67]: EMBASE, MEDLINE, Web of Science, Scopus, CINAHL, PsycINFO and Education Resources Information Center (ERIC).

Following database selection, researchers must develop a set of search terms for each. This involves two stages: initial development and agreement of search terms (typically for the MEDLINE database), and the adaptation of these for application in other databases. There is significant work involved in developing search terms, which need to be sufficiently broad to capture all potentially relevant records but precise enough that the number of returns can be screened within available resources. For initial development, it is crucial to identify the different aspects of the review question to be reflected in the search to build ‘sets’ of search terms for each (Table 2 illustrates the appropriate development of sets for our example reviews). To build the initial MEDLINE search strategy, begin the population of each ‘set’ with Medical Subject Headings (MeSH terms) via the use of MeSH Browser [68] (which returns relevant MeSH terms when keywords are entered). Following this, augment subject headings within each set with free-text keywords (i.e. relevant words or phrases that describe the concept/issue/topic/intervention). These can be generated based on researchers’ knowledge but should be supplemented with additional keywords from the search strategies of other related reviews and the consideration of

title, abstracts and keywords used within known relevant articles.

Typically, search terms within sets are combined using the 'OR' Boolean operator (the instruction of a database to find articles using this term OR that term OR this term etc.), while the sets themselves are combined using the 'AND' Boolean operator (the instruction of a database to find articles that use at least one term from Set 1 AND at least one term from Set 2). When a full search strategy is available, researchers should carefully consider the application of any search filters (e.g. language, peer-review status), as their use substantially impacts the returns generated and requires justification. It is important to ensure that the same search filter can be applied across each database.

Finally, researchers will need to add wildcards or truncation to their search strategy as appropriate. This involves adding a term or symbols to individual search terms to allow identification of different forms of those words (e.g. entering 'simulation' will return records using that term only, entering 'simulat*' will return records using simulations, simulates, simulated, etc.). This allows researchers to capture plurals (e.g. measurement and measurements), alternate spellings (e.g. behaviour and behavior) and potentially different word endings (e.g. mannequin and mannekin). Wildcards also support researchers in searching within specific parts of an article (e.g. ti,ab. entered on MEDLINE will implement searches for the relevant term only in articles' title or abstract) and searching for words only when used in proximity (e.g. 'high adj1 fidelity', 'task adj1 trainer'). The symbols for enacting truncation and wildcards differ across databases and researchers should use the search guidance available within databases to ensure an equivalent search. HS researchers may find search 'hedges' [69] – pre-defined comprehensive combinations of search terms for a concept/topic created by an expert searcher, typically for a single database, useful to save time or narrow the scope of the search.

Best practice suggests engaging a research librarian in, at minimum, review of the initial search terms, with discernible positive effects of librarian involvement evidenced [70]. The Peer Review of Electronic Search Strategies (PRESS) [37] checklist can be applied by researchers to critically appraise their search strategy. In advance of formally conducting the search, it may also be useful to first identify a 'seed set' of articles (i.e. a sample set of 4–6 articles that meet inclusion criteria and would be expected to be captured by the search strategy); these can then be used to test the thoroughness of the database searches, ensuring that search terms and processes would capture such relevant articles. Once an initial search strategy has been agreed, researchers must adapt it for use within the remaining databases. While the MeSH browser yields subject headings existing in MEDLINE and PubMed, other databases will have their own, tailored to their specific focus. Researchers can identify these terms within each database's relevant thesaurus or library.

Several issues may arise during the process of developing the search strategy. An excessive number of returns will require researchers to reconsider keywords, explore an

additional 'set' of search terms, and/or incorporate the use of additional wildcards (e.g. proximity operators). Conversely, few returns may necessitate broadening the search through removal of a 'set' of terms, introducing additional keywords, removing search filters and/or revising use of truncation or wildcards. Once the electronic database search strategy has been finalised for each database, its development is complete. The PRISMA 2020 statement [42] requires presentation of a full search strategy for each database, typically within the supplementary materials. Researchers should ensure that the search strategy for each database is saved, allowing searches to be enacted quickly and consistently and facilitating the updating of searches later. Updating searches prior to completion of the SR may be considered to ensure the most recent publications are included. There is no clear guidance on acceptable time lapse since completion of searching, although some journals require searches to have been completed within the previous six months.

Once plans for electronic searches have been agreed, researchers must decide on the use of any supplemental search tactics to identify additional relevant studies missed by electronic searches and to ensure the comprehensiveness of their overall search strategy. Table 3 provides an overview of supplemental search tactics and guidance on employing each. While the target within a SR is identifying all relevant studies, it is recognised that there will be instances where relevant articles are not identified, which is not necessarily an indicator of a poor-quality search. Where researchers have adequately planned and carefully implemented a search strategy, the likelihood is that the resulting review and its conclusions will be useful and well-founded despite this [38].

Step 5: screen and select studies

Once a researcher has retrieved relevant records from the electronic databases, the process of screening these and selecting studies for inclusion begins, comprising de-duplication, initial screening and second-stage screening.

Searching multiple databases inevitably results in the retrieval of duplicate publications (e.g. different databases returning the same publication, or different versions of the same study such as a conference abstract and a full text article) [72]. Each study should only be considered once when screening; therefore, researchers should use an electronic reference management system (e.g. EndNote, ProCite) to facilitate the identification and removal of duplicates (i.e. de-duplication) [18,72,73]. De-duplication lessens workload [72], ensuring reviewers avoid wasting time screening the same citation multiple times.

During the initial screening of remaining records' titles and abstracts, researcher(s) will evaluate each through application of the pre-defined eligibility criteria. Two researchers screening records independently is ideal to ensure transparency and minimise bias [18,42] and increases the number of relevant studies identified [74]. Each researcher screens each record independently and determines whether there is enough information in the

Table 3: Supplemental search tactics for consideration based on resources available and anticipated body of research

| Search tactic | Description and guidance |
|---------------------------------------|--|
| Reference list screening (RLS) | <p>RLS is also referred to as ancestry searching or backwards citation chasing. It involves manually screening the reference lists of all studies identified for inclusion through the electronic database searches and is typically recommended within systematic reviews. The rationale for this process is that 'like cites like' and that studies relevant for inclusion may be expected to cite other articles that have a similar focus/aim/intervention.</p> <p>RLS may also take the form of examining the reference lists of related review articles that have been identified. RLS will generally yield a small number of additional returns; identification of many additional articles should be considered a cause for concern as it suggests that the electronic searches were potentially inadequate. Review of the new articles will support the researchers in identifying search terms that were missing and should prompt consideration of whether searches should be re-run to ensure all relevant studies have been included.</p> |
| Direct contact searching | <p>Involves contacting researchers working in the field to identify relevant unpublished research, or new and/or ongoing research articles. Individuals contacted are usually the authors of studies already identified as eligible for inclusion or active researchers in the focal area.</p> <p>Direct contact searching can be disadvantageous for several reasons including its time-consuming nature, difficulty identifying current contact details for target researchers, a low response rate and the need for repeated contacts, potentially low-quality returns in the form of unpublished data and its suitability only for relatively recent research [71]. Direct contact searching is most likely to be useful within intervention-focused reviews as a means of reducing the likelihood of publication bias impacting results. However, in a field like HS with well-indexed journals, the resources and time required to enact direct contact searching are unlikely to result in a meaningful yield.</p> |
| Forwards citation tracking | <p>Also referred to as citation chasing. The process involves reviewing articles that have cited studies which have been identified as suitable for inclusion – the converse of RLS – and may be particularly useful within reviews of rapidly evolving research areas.</p> <p>Platforms that support the examination of citations of an individual study include Web of Science, Scopus and Google Scholar. Both RLS and forwards citation tracking can typically be completed by one researcher, and the number of records examined, number of full-texts screened and number of articles ultimately included from these efforts should be recorded.</p> |
| Grey literature searches | <p>Involves the use of web searches (e.g. Google Scholar) or the use of databases focused on grey literature (e.g. ProQuest) to identify relevant non-peer-reviewed materials, often within intervention-focused reviews as an attempt to avoid the over-estimation of intervention effects. Examples may include study protocols for ongoing studies, student theses, conference abstracts or technical reports.</p> <p>Challenges include a lack of best practice guidance, difficulty conducting replicable web searches, materials of low methodological/reporting quality and the resource intensive nature of searches. However, prestigious evidence synthesis organisations (e.g. Cochrane, Campbell Collaboration) recommend its use.</p> |
| Hand searching | <p>Involves manually reviewing the records published within specific journals considered relevant to the topic. Researchers may consider this appropriate when most relevant outputs are published within a core set of journals or where core journals are not indexed within the electronic databases used.</p> <p>However, it should be noted that although HS-specific journals exist, many of the most influential simulation articles have been published in other outlets, suggesting a limited utility of hand searching within HS-focused reviews. A low precision, and lengthy process, have been identified as deterrents to its use.</p> |

Note: The guidance within this table has been informed by the work of Cooper et al. [71].

title and abstract to include or exclude or whether it is potentially relevant and requires further review. Next, researchers compare assessments; records should be excluded if both reviewers agree, and records should be moved to the next stage (i.e. full-text screening) if both reviewers indicate eligibility or potential eligibility. Records with differing judgements should be reviewed together, and consensus achieved via discussion [74]. Those articles identified as relevant, or potentially relevant, should be obtained for second-stage screening [73]. In practice, engaging two researchers in study screening may not be feasible and will require consideration of capacity and resources available. Using a second reviewer during full-text screening only may be a way to reduce resources, while still limiting bias [74].

Second-stage screening involves examination of the full text of records retained from title and abstract screening,

with involvement of two researchers strongly recommended [74]. Each researcher should scrutinise the entire text of each record to confirm/disconfirm relevance, guided by the eligibility criteria, and decide whether to 'include' or 'exclude' each article. For each excluded article, a justification should be documented [72]. When independent screening is complete, the researchers' logs should be compared in the same manner described previously. The involvement of a third researcher, or wider research team, is important for resolving any remaining disagreements or inconsistencies. Ultimately, a final list of included studies from electronic database searches will have been identified [18]. The use of supplemental search tactics (e.g. reference list reviews) can be completed at this point (see Table 3).

Adequate documentation of screening (i.e. a record of the count of records identified, screened, included, excluded and reasons) is essential [72] for completing the PRISMA

flowchart [18]. We direct readers to Rethlefen and Page [75] for further useful guidance on tracking records throughout the SR process.

Step 6: extract data and finalise a record of studies

When the studies for inclusion have been finalised, data extraction can begin. Data extraction refers to the process of recording relevant information from studies for consideration, refinement and analysis [76]. The data extraction process comprises selecting data collection tools, constructing data collection forms and abstracting, managing and archiving data [77].

Data extraction can be conducted using either paper forms, electronic forms (e.g. Microsoft Word), spreadsheet software (e.g. Microsoft Excel) or commercial data systems (e.g. Covidence) that allow online form building, data entry by several users, data sharing and efficient data management [77]. Electronic forms are more efficient and less error-prone than paper forms, while data systems require greater financial investment and training [77,78]. However, no single data-extraction tool is best for all SRs in all circumstances. Therefore, researchers should carefully consider the volume, nature and anticipated complexity of the data and resource availability, before selecting a tool [78]. We centre our subsequent guidance on using an electronic form via Microsoft Word, which is generalisable to other tools.

The data extraction form should be easy to use, standardised and allow for the structured collation of data to sufficiently represent the study [77], minimising the need to return to the source [79]. Careful development, pilot testing and application are crucial, and should involve iterative input from all researchers. First, researcher(s) should prepare outlines of the tables and figures for the SR (e.g. table of study characteristics, tables/figures offering a summary of the findings) to support determination of the data required [77,79]. Next, data extraction variables/elements should be logically grouped (see Table 2 for illustration of data extraction for example reviews). Subsequently, the form can be presented using a word processor, in a table format capturing all studies or using individual forms that can later be tabulated. A codebook should also be generated, containing relevant definitions and details on how to capture and code data for each variable. Providing guidance on the form itself supports quality and consistency across researchers [77]. Finally, the form should be piloted among multiple researchers, extracting data from at least three articles [77], with modifications made as necessary to ensure agreement. The 'seed set' of articles developed within Step 4 for trialling the search strategy, could also be used within the pilot data extraction.

It is imperative that researchers (especially SR novices) receive training to familiarise them with the data extraction process and clarify any uncertainties. Data extraction errors are frequent [80]; therefore, two researchers should be involved [4,77], which has been found to result in fewer errors [81]. Data extractors from complementary disciplines (e.g. a methodologist and subject-matter expert) [77,79]

may valuably support one another. Each researcher should extract data for all studies independently, then compare their records to identify discrepancies. Discrepancies can be resolved through discussion, with re-engagement with the source document as necessary. If consensus cannot be achieved, a third researcher may adjudicate. A record of the presence and resolution of disagreements, and raw copies of original individual data extraction forms or files (this will differ based on whether the researchers have opted to extract data from all studies into one table or used individual forms), should be retained [79]. Ultimately, files should be merged into one final single document that offers a complete and agreed record of data extraction.

To facilitate the later synthesis of extracted data, researcher(s) may have to perform a degree of coding when initial data extraction is completed, introducing homogeneity within the data extraction records [82]. This can be achieved via content analysis, which involves researchers developing categories/codes to capture the extracted data either *de novo* (i.e. inductive coding [83]) or through the application of existing frameworks to standardise the data (i.e. deductive coding [83]). Such coding allows researchers to establish the frequency and consistency of findings, methods or study characteristics and helps to determine what is 'key' across studies or how they reflect existing theory or understanding.

Step 7: critically appraise studies

Given the implications of SR for practice and research, consideration of trustworthiness via critical appraisal of included studies is core to conducting a high-quality SR [42]. Researchers will encounter two relevant concepts:

- Risk of bias: how methodological limitations or flaws within design, conduct or analysis may compromise resulting data [84]; and
- Study/methodological quality/rigour: how closely studies resemble best practice in conduct and design [84].

Focusing on risk of bias is recommended [84], particularly within *effectiveness* reviews. Several tools and corresponding guidance on their use, have been developed to support researchers in formally appraising studies. Our non-systematic examination of recent HS-related SRs revealed a great number of tools have been used to support critical appraisal, including:

- ROB-2 (Cochrane's risk-of-bias tool for randomised trials) [85];
- Mixed Methods Assessment Tool (MMAT) [86];
- COSMIN Risk-of-Bias tool to assess the quality of studies on measurement properties [34];
- JBI's range of critical appraisal tools [87];
- Critical Appraisal Skills Programme (CASP) checklists [88];
- Medical Education Research Study Quality Instrument (MERSQI) [39]; and
- Risk of Bias in Non-randomized Studies of Interventions tool [40].

When selecting a tool, researchers should consider the likely study designs, and the experience within the research team.

Considering critical appraisal tools used in similar SRs, consulting colleagues and experts in evidence synthesis, and examining other review articles that have used a tool under consideration may support decision-making.

Typically, two researchers will independently complete critical appraisal with discussion to agree final ratings, though this process is sometimes conducted in tandem, with researchers working together to finalise judgements for each item and study. Numerous challenges exist within the appraisal process: considerable time is required [89], inter-rater agreement is variable even with well-established tools [89], and performance is likely impacted by familiarity with the tool, between co-raters and with critical appraisal [89]. Accordingly, the tool’s user-friendliness, sufficient allocation of time, researcher experience and initial training, review and piloting of the tool are crucial considerations to ensure the process is completed effectively.

Step 8: synthesise and write-up

When the data extraction record is finalised, the process of data synthesis commences. SRs in HS frequently emphasise diversity across included studies (e.g. heterogeneity of outcome measures [27,28]) which is likely to render synthesis via meta-analysis (see Table 1) challenging. Accordingly, within both *effectiveness reviews* and other review types, a range of alternative approaches are often more suitable. We focus our guidance herein on the process of narrative synthesis, which we expect to be suited to most reviews of the HS literature, though Table 4 details alternative tactics for synthesising data in SRs that can also be applied.

Narrative synthesis is a textual approach to collating findings, involving the use of text to ‘tell the story’ of the

findings and learning [16]. Narrative synthesis is inherently subjective to an extent [90], and it is important that researchers pursue a rigorous and transparent approach to reduce potential bias. Popay et al. [16] offer comprehensive guidance on the conduct and reporting of narrative synthesis, defining four main elements. First, developing a theory about why the intervention works and for whom. This element is applicable to *effectiveness* SRs and potentially others such as *experiential or qualitative* SRs. In addressing this, researchers consider the rationale for the intervention outlined in studies seeking to identify the pathways through which an intervention has its effects [91]. Researchers focusing on the effects of a HS intervention will need to organise the results of included studies to enable description of emerging patterns in terms of the direction and size of effects [16]. Second, researchers must develop a preliminary synthesis of findings of included studies. This involves determining how best to bring together, organise and describe the results from included studies. Ryan [91] offers valuable guidance on varied ways to achieve this:

- Textual descriptions summarising features for each study;
- Grouping the studies (e.g. by intervention, population groups, study design, etc.);
- Tabulating results across the included studies to identify patterns;
- Transforming the data (e.g. into a common statistical or descriptive format; see description of coding in Step 6);
- Using vote counting to provide an initial description of results – typically acceptable only as an interim step in assessing data;
- Translating data using thematic or content analysis to identify recurrent concepts or themes.

Table 4: Supplemental or alternative tactics for consideration based on focus of review and nature of anticipated body of research

| Synthesis tactic | Description and guidance |
|---|---|
| Summarising effect estimates | involves using descriptive statistics (e.g. using median, interquartile range, range) to provide information on the range and distribution of observed effects and is typically used to synthesise results in cases where it is difficult to undertake a meta-analysis [93]. McKenzie and Brennan [93] provide guidance on how best to complement reporting of the summary statistics with visual displays. Limitations include not accounting for differences in the relative sizes of the studies, and lack of evaluation of the performance of these statistics applied in the context of summarising effect estimates [93]. |
| Combining p-values | (i.e. significance levels) of each study involves conducting a statistical test to determine whether there is evidence that the intervention is effective in at least one study. These can be used to synthesise results when included studies report minimal information beyond p-values and direction of effect, results of non-parametric tests or results of heterogeneous outcomes and statistical tests [93]. Several methods exist, of which the most commonly applied is Fisher’s method [94] which combines the p-values from statistical tests across studies using a formula. Researchers should consult guidance from Cooper et al. [94] on applying Fisher’s method. Limitations include the lack of information on the magnitude of effects, and difficulty in interpreting the test results when statistically significant [93]. |
| Vote counting based on direction of effect | involves comparing the number of effects showing benefit to the number of effects showing harm for a particular outcome based on the observed direction of effect alone (regardless of statistical significance), thereby creating a standardised binary metric [93,95]. This can be used to synthesise results when only direction of effect is reported, or there is inconsistency in the effect measures or data reported across studies [93]. Conventional forms of vote counting based on the count on the statistical significance and direction of the effect estimates, which has considerable limitations and should be avoided [93,96]. |

Note: The guidance within this table has been informed by the work of McKenzie and Brennan [93].

Third, researchers must explore relationships in the data within and between studies. Researchers should explore two key relationships: those between characteristics of individual studies and their findings, and those between the findings of different studies [16]. To achieve this, the researcher should examine the extracted data with a view to looking at the relationships between study results and key aspects of the studies and comparing these relationships across studies. At this point, the researcher should consider whether variability in study design, populations, interventions or outcomes is likely to explain these differences, in addition to the effect of heterogeneity [91]. The following approaches to explore the relationships between and within studies can be used [91]:

- Visual tools (e.g. forest plots, graphs) can help to organise and present results. These should supplement, rather than replace, a textual description of results.
- Subgroup analysis can be used to explore the impact of potential effect modifiers on intervention effects.
- Developing conceptual models (e.g. idea webbing, concept mapping).

Finally, researchers should appraise the rigour of the synthesis. This comprises examining the quality or risk of bias of included studies (see *Step 7*). Researchers should also consider the trustworthiness of their processes for collating evidence, such as critical reflection on strengths and weaknesses of the methods used to narratively synthesise included studies [91]. When reporting on the narrative synthesis of studies, researchers should adhere to the Synthesis Without Meta-analysis (SWiM) guideline [92] to ensure a clear description of their procedures.

The remaining work of the SR consists of preparing a written manuscript. There is significant value in identifying a target journal prior to commencing the write-up, as this will guide certain elements (e.g. structure, word count etc.). Considering other SRs published in the target journal may also support authors in preparing a suitable manuscript. A SR manuscript will include the following sections: Introduction, Methods, Results and Discussion along with an Abstract summarising the conduct and core messages of the SR.

Drawing conclusions in a SR typically involves considering the implications for practice, policy and future research. When providing implications for theory, practice or policy, researchers should address the review question and consider the intended end-users' needs. These should be supported by the evidence synthesis (i.e. they should not be speculative), with acknowledgement of the strength and limitations of the evidence [45,97]. In stating recommendations for future research, researchers should avoid vague statements (e.g. 'more research is needed') and identify specific areas where evidence is lacking and warrants prioritisation [45]. The PICO domains may be useful here to inform aspects that could be addressed more effectively in the future [97]. Further, the domains of the GRADE [98] certainty framework can support researchers in interpreting their evidence and drawing conclusions about future research and practice.

Researchers must carefully consult and adhere to the PRISMA 2020 checklist [42], which offers detailed instruction for writing each manuscript component and for appropriately detailing funding, conflicts of interest and author contributions. This guidance includes 12 items that dictate good practice in the presentation of an Abstract, the preparation of which is often most appropriate when the main manuscript text has been completed. Writing within the confines of limited word counts is challenging, and the considered use of tables and figures, as well as supplementary materials (e.g. for detailed study-by-study summaries, full search strategies), is required to ensure that the SR's data is presented comprehensively and effectively [99].

Considerations for conducting a systematic review

A variety of resources exist that may support completion of certain SR processes more efficiently and facilitate more effective collaboration. There exists an array of high-quality training programmes that researchers are encouraged to engage with to support the delivery of rigorous SRs. While 'how-to' guides such as that presented herein, or indeed dedicated textbooks or guides, such as the JBI Manual for Evidence Synthesis [100], offer significant value, they cannot replace in-person training that can address specific researcher queries. Software to support completion of the SR process includes RevMan, Covidence and Rayyan, with access often requiring a paid subscription. Recent research has suggested automation tools, often freely available, save time without a loss of methodological quality, as compared to novice researchers [101]. Full coverage of SR automation tools and software is beyond the scope of this guide (see Affengruber and colleagues [102] for more detail), but some examples include:

- Forbes and colleagues [103] describe 'the Deduplicator', an automation tool for removing duplicate records. The Deduplicator tool reduced the time required to remove duplicate records, with at least the same if not better, accuracy than typical semi-manual deduplication processes.
- Chai et al. [104] evaluated a semi-automated tool 'Research Screener' that supports screening study abstracts to identify articles for inclusion. The authors estimated a time saving of over 12.5 days compared to manual screening.
- Motzfeldt Jensen and colleagues [105] demonstrated that ChatGPT-4o could reliably duplicate data extraction (92.4% accurate).

While generally positive attitudes towards automation tools exist, their use is limited by a lack of knowledge or training to support researchers in appropriately employing them [106]. This is important, as the output of these tools may appear valid but requires careful review by researchers with sufficient knowledge to identify errors [107].

Next, researchers should be aware of reporting guidelines relevant to the body of literature being synthesised and of the potential value of these for supporting the conduct of a SR. These include reporting guidelines specific to HS (e.g. Standards for Quality Improvement Reporting Excellence for SIMulation [SQUIRE-SIM] [108]; Reporting Guidelines for Health Care Simulation Research [109]) as well as those relating to the research design of included studies (e.g.

CONsolidated Standards Of Reporting Trials; [CONSORT] [110] or Strengthening the Reporting of Observational Studies in Epidemiology [STROBE] [111]). A consistent challenge in HS scholarship has been the insufficient description of the simulation intervention itself [109]. Accordingly, application of reporting guidelines within a SR can help clarify the extent to which the use of HS can be understood and replicated, supporting appraisal of this element of studies and clarifying where uncertainty about how HS is being used remains and where attention should be directed in future research. Reporting guidelines may also be used to facilitate the screening and selection of studies, reducing a body of literature within an SR to only high-quality studies that offer the information crucial to understanding how HS has been used and what its effects might be.

Finally, researcher awareness of the research context regarding reviews of HS research literature is important. SRs are accepted for publication in the core HS-specific journals, including *the Journal of Healthcare Simulation*, *Simulation in Healthcare*, *Advances in Simulation*, *Clinical Simulation in Nursing* and *the Journal of Surgical Simulation*. However, the publication of HS research and reviews in other journals (e.g. medical specialty journals, medical education journals) is certainly possible; indeed, many of the most highly cited HS research articles have been published in such journals [6]. The potential to secure research funding to support the conduct of a SR in HS will vary depending on a researcher's location and role. This may be challenging, as medical education research is consistently under-funded [112]. Kunkler [113] and Gruppen and Durning [112] offer useful direction on securing research funding. However, researchers seeking to contribute to the advancement of HS should not let funding challenges dissuade them from completing important SRs. SRs may be highly impactful, being frequently among the most highly cited HS articles [6], and informing core developments in the field (e.g. serving as the foundation for The International Nursing Association for Clinical Simulation and Learning's 'Healthcare simulation standards of best practice' [114]).

Conclusion

The work of conducting a methodologically rigorous SR is considerable. However, SRs offer a crucial means of establishing best practice in designing, delivering and assessing applications of HS that can be disseminated widely and charting the paths that will allow HS research and practice to evolve and strengthen. We hope this guide will support researchers' success across all steps of the SR process and that the field will benefit accordingly.

Suggestions for further reading

- Higgins JP, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Cochrane and John Wiley & Sons Ltd. 2008.
- Affengruber L, van der Maten MM, Spiero I, Nussbaumer-Streit B, Mahmić-Kaknjó M, Ellen ME, et al. An exploration of available methods and tools to improve the efficiency of systematic review production: a scoping review. *BMC Medical Research Methodology*. 2024;24(1), 210. doi: [10.1186/s12874-024-02320-4](https://doi.org/10.1186/s12874-024-02320-4)

- Kunkler K. Identifying and applying for funding. In: Nestel D, Hui J, Kunkler K, Scerbo MW, Calhoun AW, editors. *Healthcare simulation research: a practical guide*. Springer. 2019, p.269–276.
- Aromataris ELC, Porritt K, Pilla B, Jordan Z. *JBIC manual for evidence synthesis* [Internet]. JBI. 2024. Available from: <https://jbi-global-wiki.refined.site/space/MANUAL> [Accessed 20 September 2025].
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372. doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)
- Tacconelli E. Systematic reviews: CRD's guidance for undertaking reviews in health care. *The Lancet Infectious Diseases*. 2010;10(4):226.
- Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. *Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC methods programme Version*. 2006;1(1):b92.

Acknowledgements

The Association for Simulated Practice in Healthcare (ASPIH) has supported the publication of this work through its fee waiver member benefit.

Declarations

Authors' contributions

SL and CM contributed equally to the planning, development, writing, and revisions of this article.

Funding

None declared.

Availability of data and materials

None declared.

Ethics approval and consent to participate

Not applicable.

Competing interests

None declared.

References

1. Gough D, Davies P, Jamtvedt G, Langlois E, Littell J, Lotfi T, et al. Evidence synthesis international (ESI): position statement. *Systematic Reviews*. 2020;9(1):155. doi: [10.1186/s13643-020-01415-5](https://doi.org/10.1186/s13643-020-01415-5)
2. Colby DC, Quinn BC, Williams CH, Bilheimer LT, Goodell S. Research glut and information famine: making research evidence more useful for policymakers. *Health Affairs*. 2008;27(4):1177–1182. doi: [10.1377/hlthaff.27.4.1177](https://doi.org/10.1377/hlthaff.27.4.1177)
3. Karlsson LE, Takahashi R. A resource for developing an evidence synthesis report for policy-making [Internet]. Copenhagen: WHO Regional Office for Europe. 2017. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK453541/> [Accessed 20 September 2025].
4. Higgins JP, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Cochrane and John Wiley & Sons Ltd. 2008.

5. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*. 2018;18(1):143. doi: [10.1186/s12874-018-0611-x](https://doi.org/10.1186/s12874-018-0611-x)
6. Walsh C, Lydon S, Byrne D, Madden C, Fox S, O'Connor P. The 100 most cited articles on healthcare simulation: a bibliometric review. *Simulation in Healthcare*. 2018;13(3):211–220. doi: [10.1097/SIH.0000000000000293](https://doi.org/10.1097/SIH.0000000000000293)
7. Essex R, Weldon SM, Markowski M, Gurnett P, Slee R, Cleaver K, et al. A systematic mapping literature review of ethics in healthcare simulation and its methodological feasibility. *Clinical Simulation in Nursing*. 2022;73:48–58. doi: [10.1016/j.ecns.2022.07.001](https://doi.org/10.1016/j.ecns.2022.07.001)
8. Decker S, Sapp A, Bibin L, Brown MR, Chidume T, Crawford SB, et al. The impact of simulation debriefing process on learning outcomes: an umbrella review. *Clinical Simulation in Nursing*. 2025;101:101715. doi: [10.1016/j.ecns.2025.101715](https://doi.org/10.1016/j.ecns.2025.101715)
9. Slavinska A, Palkova K, Grigoroviča E, Edelmers E, Pētersons A. Narrative review of legal aspects in the integration of simulation-based education into medical and healthcare curricula. *Laws*. 2024;13(2):15. doi: [10.3390/laws13020015](https://doi.org/10.3390/laws13020015)
10. Denniston C, Molloy E, Nestel D, Woodward-Kron R, Keating JL. Learning outcomes for communication skills across the health professions: a systematic literature review and qualitative synthesis. *BMJ Open*. 2017;7(4):e014570. doi: [10.1136/bmjopen-2016-014570](https://doi.org/10.1136/bmjopen-2016-014570)
11. Seabrooke M, Seabrooke A. In situ clinical education of frontline healthcare providers in under-resourced areas: a rapid review. *Canadian Journal of Rural Medicine*. 2024;29(1):20–29. doi: [10.4103/cjrm.cjrm_95_22](https://doi.org/10.4103/cjrm.cjrm_95_22)
12. Smallheer B, Chidume T, M'lyn K, Dawkins D, Pestano-Harte M. A scoping review of the priority of diversity, inclusion, and equity in health care simulation. *Clinical Simulation in Nursing*. 2022;71:41–64. doi: [10.1016/j.ecns.2022.05.009](https://doi.org/10.1016/j.ecns.2022.05.009)
13. Mahdi SS, Jafri HA, Allana R, Battineni G, Khawaja M, Sakina S, et al. Systematic review on the current state of disaster preparation Simulation Exercises (SimEx). *BMC Emergency Medicine*. 2023;23(1):52. doi: [10.1186/s12873-023-00824-8](https://doi.org/10.1186/s12873-023-00824-8)
14. Mazzone E, Puliatti S, Amato M, Bunting B, Rocco B, Montorsi F, et al. A systematic review and meta-analysis on the impact of proficiency-based progression simulation training on performance outcomes. *Annals of Surgery*. 2021;274(2):281–289. doi: [10.1097/SLA.0000000000004650](https://doi.org/10.1097/SLA.0000000000004650)
15. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*. 2009;26(2):91–108. doi: [10.1111/j.1471-1842.2009.00848.x](https://doi.org/10.1111/j.1471-1842.2009.00848.x)
16. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme. Version 1. Lancaster, UK: University of Lancaster. 2006.
17. Munn Z, Porritt K, Lockwood C, Aromataris E, Pearson A. Establishing confidence in the output of qualitative research synthesis: the ConQual approach. *BMC Medical Research Methodology*. 2014;14(1):108. doi: [10.1186/1471-2288-14-108](https://doi.org/10.1186/1471-2288-14-108)
18. Pollock A, Berge E. How to do a systematic review. *International Journal of Stroke*. 2018;13(2):138–156. doi: [10.1177/1747493017743796](https://doi.org/10.1177/1747493017743796)
19. Kolaski K, Romeiser Logan L, Ioannidis JP. Guidance to best tools and practices for systematic reviews. *Journal of Pediatric Rehabilitation Medicine*. 2023;16(2):241–273. doi: [10.3233/PRM-230019](https://doi.org/10.3233/PRM-230019)
20. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Medical Research Methodology*. 2018;18(1):5. doi: [10.1186/s12874-017-0468-4](https://doi.org/10.1186/s12874-017-0468-4)
21. O'Connor P, O'Dea A, Byrne D, editors. Introduction, history and key concepts [Internet]. In: *The essential handbook of healthcare simulation* [Internet]. Boca Raton: CRC Press. 2023. p. 1. Available from: <https://www-taylorfrancis-com.nuigalway.idm.oclc.org/books/mono/10.1201/9781003296942/essential-handbook-healthcare-simulation-paul-connor-dara-byrne-angela-dea> [Accessed 5 September 2025].
22. Weldon SM, BATTERY AG, Spearpoint K, Kneebone R. Transformative forms of simulation in health care—the seven simulation-based 'I's: a concept taxonomy review of the literature. *International Journal of Healthcare Simulation*. 2023:1–13. doi: [10.54531/tzfd6375](https://doi.org/10.54531/tzfd6375)
23. O'Connor P, O'Dea A, Byrne D, editors. Running a simulation facility. In: *The essential handbook of healthcare simulation* [Internet]. Boca Raton: CRC Press. 2023. p. 92. Available from: <https://www-taylorfrancis-com.nuigalway.idm.oclc.org/books/mono/10.1201/9781003296942/essential-handbook-healthcare-simulation-paul-connor-dara-byrne-angela-dea> [Accessed 5 September 2025].
24. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. Effect of practice on standardised learning outcomes in simulation-based medical education. *Medical Education*. 2006;40(8):792–797. doi: [10.1111/j.1365-2929.2006.02528.x](https://doi.org/10.1111/j.1365-2929.2006.02528.x)
25. McGaghie WC, Draycott TJ, Dunn WF, Lopez CM, Stefanidis D. Evaluating the impact of simulation on translational patient outcomes. *Simulation in Healthcare*. 2011;6(7):S42–S47. doi: [10.1097/SIH.0b013e318222fde9](https://doi.org/10.1097/SIH.0b013e318222fde9)
26. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA*. 2011;306(9):978–988. doi: [10.1001/jama.2011.1234](https://doi.org/10.1001/jama.2011.1234)
27. Yousef KM, Alananzeh I, Beegom S, Chavez J, Hatahet S, Khalil H, et al. Assessing outcome measurements and impact of simulation in neurocritical care training: a systematic review. *Journal of Neuroscience Nursing*. 2024;56(4):130–135. doi: [10.1097/JNN.0000000000000767](https://doi.org/10.1097/JNN.0000000000000767)
28. Jackson M, McTier L, Brooks LA, Wynne R. Impact of simulation design elements on undergraduate nursing education: a systematic review. *Clinical Simulation in Nursing*. 2024;89:101519. doi: [10.1016/j.ecns.2024.101519](https://doi.org/10.1016/j.ecns.2024.101519)
29. Møller A, Myles P. What makes a good systematic review and meta-analysis? *British Journal of Anaesthesia*. 2016;117(4):428–430. doi: [10.1093/bja/aew264](https://doi.org/10.1093/bja/aew264)
30. Stuart D, Kennedy K. Right review: a web-based tool for review-based research. *Journal of Electronic Resources in Medical Libraries*. 2025;22(1–2):1–8. doi: [10.1080/15424065.2024.2423942](https://doi.org/10.1080/15424065.2024.2423942)
31. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO

- registry. *BMJ Open*. 2017;7(2):e012545. doi: [10.1136/bmjopen-2016-012545](https://doi.org/10.1136/bmjopen-2016-012545)
32. Hulley S, Cummings S, Browner W, Grady D, Newman T. *Designing clinical research*. 4th edition. Philadelphia, PA: Lippincott Williams and Wilkins. 2013.
 33. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP Journal Club*. 1995;123(3):A12–A13. PMID: 7582737.
 34. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research*. 2010;19(4):539–549. doi: [10.1007/s11136-010-9606-8](https://doi.org/10.1007/s11136-010-9606-8)
 35. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, De Vet HC, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*. 2018;27(5):1147–1157. doi: [10.1007/s11136-018-1798-3](https://doi.org/10.1007/s11136-018-1798-3)
 36. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*. 2015;4(1):1. doi: [10.1186/2046-4053-4-1](https://doi.org/10.1186/2046-4053-4-1)
 37. Blackwood D. Peer Review of Electronic Search Strategies (PRESS): ‘Can you check my systematic review search strategy?’. *HLA News*. 2015;9–10. doi: [10.3316/informit.314342355541565](https://doi.org/10.3316/informit.314342355541565)
 38. Atkinson KM, Koenka AC, Sanchez CE, Moshontz H, Cooper H. Reporting standards for literature searches and report inclusion criteria: making research syntheses more transparent and easy to replicate. *Research Synthesis Methods*. 2015;6(1):87–95. doi: [10.1002/jrsm.1127](https://doi.org/10.1002/jrsm.1127)
 39. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. *JAMA*. 2007;298(9):1002–1009. doi: [10.1001/jama.298.9.1002](https://doi.org/10.1001/jama.298.9.1002)
 40. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355: i4919. doi: [10.1136/bmj.i4919](https://doi.org/10.1136/bmj.i4919)
 41. Harrison R, Jones B, Gardner P, Lawton R. Quality assessment with diverse studies (QuADS): an appraisal tool for methodological and reporting quality in systematic reviews of mixed-or multi-method studies. *BMC Health Services Research*. 2021;21(1):1–20. doi: [10.1186/s12913-021-06122-y](https://doi.org/10.1186/s12913-021-06122-y)
 42. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)
 43. Thabane L, Thomas T, Ye C, Paul J. Posing the research question: not so simple. *Canadian Journal of Anesthesia*. 2009;56(1):71–79. doi: [10.1007/s12630-008-9007-4](https://doi.org/10.1007/s12630-008-9007-4)
 44. Stone P. Deciding upon and refining a research question. *Palliative Medicine*. 2002;16(3):265–267. doi: [10.1191/0269216302pm562xx](https://doi.org/10.1191/0269216302pm562xx)
 45. Tacconelli E. Systematic reviews: CRD’s guidance for undertaking reviews in health care. *The Lancet Infectious Diseases*. 2010;10(4):226. doi: [10.1016/S1473-3099\(10\)70065-7](https://doi.org/10.1016/S1473-3099(10)70065-7)
 46. Riva JJ, Malik KM, Burnie SJ, Endicott AR, Busse JW. What is your research question? An introduction to the PICOT format for clinicians. *The Journal of the Canadian Chiropractic Association*. 2012;56(3):167. PMID: 22997465
 47. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research*. 2012;22(10):1435–1443. doi: [10.1177/1049732312452938](https://doi.org/10.1177/1049732312452938)
 48. Gomersall JS, Jadotte YT, Xue Y, Lockwood S, Riddle D, Preda A. Conducting systematic reviews of economic evaluations. *JBI Evidence Implementation*. 2015;13(3):170–178. doi: [10.1097/XEB.0000000000000063](https://doi.org/10.1097/XEB.0000000000000063)
 49. Wildridge V, Bell L. How CLIP became ECLIPSE: a mnemonic to assist in searching for health policy/management information. *Health Information & Libraries Journal*. 2002;19(2):113–115. doi: [10.1046/j.1471-1842.2002.00378.x](https://doi.org/10.1046/j.1471-1842.2002.00378.x)
 50. Harris JD, Quatman CE, Manring MM, Siston RA, Flanigan DC. How to write a systematic review. *The American Journal of Sports Medicine*. 2014;42(11):2761–2768. doi: [10.1177/0363546513497567](https://doi.org/10.1177/0363546513497567)
 51. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making*. 2007;7(1):16. doi: [10.1186/1472-6947-7-16](https://doi.org/10.1186/1472-6947-7-16)
 52. Weir MC, Grimshaw JM, Mayhew A, Fergusson D. Decisions about lumping vs. splitting of the scope of systematic reviews of complex interventions are not well justified: a case study in systematic reviews of health care professional reminders. *Journal of Clinical Epidemiology*. 2012;65(7):756–763. doi: [10.1016/j.jclinepi.2011.12.012](https://doi.org/10.1016/j.jclinepi.2011.12.012)
 53. Patino CM, Ferreira JC. Inclusion and exclusion criteria in research studies: definitions and why they matter. *Jornal Brasileiro de Pneumologia*. 2018;44:84. doi: [10.1590/s1806-37562018000000088](https://doi.org/10.1590/s1806-37562018000000088)
 54. Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*. 2003;96(3):118–121. doi: [10.1258/jrsm.96.3.118](https://doi.org/10.1258/jrsm.96.3.118)
 55. Stewart L, Moher D, Shekelle P. Why prospective registration of systematic reviews makes sense. *Systematic Reviews*. 2012;1(1):7. doi: [10.1186/2046-4053-1-7](https://doi.org/10.1186/2046-4053-1-7)
 56. Ge L, Tian JH, Li YN, Pan JX, Li G, Wei D, et al. Association between prospective registration and overall reporting and methodological quality of systematic reviews: a meta-epidemiological study. *Journal of Clinical Epidemiology*. 2018;93:45–55. doi: [10.1016/j.jclinepi.2017.10.012](https://doi.org/10.1016/j.jclinepi.2017.10.012)
 57. van der Braak K, Ghannad M, Orelia C, Heus P, Damen JA, Spijker R, et al. The score after 10 years of registration of systematic review protocols. *Systematic Reviews*. 2022;11(1):191. doi: [10.1186/s13643-022-02053-9](https://doi.org/10.1186/s13643-022-02053-9)
 58. Chang SM, Slutsky J. Debunking myths of protocol registration. *Systematic Reviews*. 2012;1(1):1–2. doi: [10.1186/2046-4053-1-4](https://doi.org/10.1186/2046-4053-1-4)
 59. Pieper D, Rombey T. Where to prospectively register a systematic review. *Systematic Reviews*. 2022;11(1):8. doi: [10.1186/s13643-021-01877-1](https://doi.org/10.1186/s13643-021-01877-1)
 60. National Institute for Health and Care Research. PROSPERO: International prospective register of systematic reviews [Internet]. Centre for Reviews and Dissemination, University of York. Available from: <https://www.crd.york.ac.uk/prospero/> [Accessed 10 September 2025].
 61. Research Registry. Research Registry [Internet]. Available from: <https://www.researchregistry.com/> [Accessed 10 September 2025].
 62. INPLASY: International Platform of Registered Systematic Review and Meta-analysis Protocols. The international

- database to register evidence synthesis projects [Internet]. Available from: <https://inplasy.com/> [Accessed 10 September 2025].
63. Centre for Open Science. Future proof your research: preregister your next study [Internet]. Available from: <https://www.cos.io/initiatives/prereg> [Accessed 10 September 2025].
 64. Protocols.io. Bring structure to your research [Internet]. Available from: <https://www.protocols.io/> [Accessed 10 September 2025].
 65. Tawfik GM, Giang HTN, Ghozy S, Altibi AM, Kandil H, Le HH, et al. Protocol registration issues of systematic review and meta-analysis studies: a survey of global researchers. *BMC Medical Research Methodology*. 2020;20(1):213. doi: [10.1186/s12874-020-01094-9](https://doi.org/10.1186/s12874-020-01094-9)
 66. Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. The contribution of databases to the results of systematic reviews: a cross-sectional study. *BMC Medical Research Methodology*. 2016;16(1):127. doi: [10.1186/s12874-016-0232-1](https://doi.org/10.1186/s12874-016-0232-1)
 67. Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Systematic Reviews*. 2017;6(1):245. doi: [10.1186/s13643-017-0644-y](https://doi.org/10.1186/s13643-017-0644-y)
 68. National Library of Medicine. Medical subject headings 2025 [Internet]. Available from: <https://meshb.nlm.nih.gov/> [Accessed 15 September 2025].
 69. Campbell S. What is the difference between a filter and a hedge? *Journal of EAHIL*. 2016;12(1):4–5.
 70. Koffel JB. Use of recommended search strategies in systematic reviews and the impact of librarian involvement: a cross-sectional survey of recent authors. *PLoS One*. 2015;10(5):e0125931. doi: [10.1371/journal.pone.0125931](https://doi.org/10.1371/journal.pone.0125931)
 71. Cooper C, Booth A, Britten N, Garside R. A comparison of results of empirical studies of supplementary search techniques and recommendations in review methodology handbooks: a methodological review. *Systematic Reviews*. 2017;6(1):234. doi: [10.1186/s13643-017-0625-1](https://doi.org/10.1186/s13643-017-0625-1)
 72. Martinez EC, Valdés JRF, Castillo JL, Castillo JV, Montecino RM, Jimenez JE, et al. Ten steps to conduct a systematic review. *Cureus*. 2023;15(12):e51422. doi: [10.7759/cureus.51422](https://doi.org/10.7759/cureus.51422)
 73. Wright RW, Brand RA, Dunn W, Spindler KP. How to write a systematic review. *Clinical Orthopaedics and Related Research*. 2007;455:23–29. doi: [10.1097/BLO.0b013e31802c9098](https://doi.org/10.1097/BLO.0b013e31802c9098)
 74. Stoll CR, Izadi S, Fowler S, Green P, Suls J, Colditz GA. The value of a second reviewer for study selection in systematic reviews. *Research Synthesis Methods*. 2019;10(4):539–545. doi: [10.1002/jrsm.1369](https://doi.org/10.1002/jrsm.1369)
 75. Rethlefsen ML, Page MJ. PRISMA 2020 and PRISMA-S: common questions on tracking records and the flow diagram. *Journal of the Medical Library Association*. 2022;110(2):253. doi: [10.5195/jmla.2022.1449](https://doi.org/10.5195/jmla.2022.1449)
 76. Büchter RB, Rombey T, Mathes T, Khalil H, Lunny C, Pollock D, et al. Systematic reviewers used various approaches to data extraction and expressed several research needs: a survey. *Journal of Clinical Epidemiology*. 2023;159:214–224. doi: [10.1016/j.jclinepi.2023.05.027](https://doi.org/10.1016/j.jclinepi.2023.05.027)
 77. Li T, Vedula SS, Hadar N, Parkin C, Lau J, Dickersin K. Innovations in data collection, management, and archiving for systematic reviews. *Annals of Internal Medicine*. 2015;162(4):287–294. doi: [10.7326/M14-1603](https://doi.org/10.7326/M14-1603)
 78. Elamin MB, Flynn DN, Bassler D, Briel M, Alonso-Coello P, Karanicolas PJ, et al. Choice of data extraction tools for systematic reviews depends on resources and review complexity. *Journal of Clinical Epidemiology*. 2009;62(5):506–510. doi: [10.1016/j.jclinepi.2008.10.016](https://doi.org/10.1016/j.jclinepi.2008.10.016)
 79. Li T, Higgins J, Deeks J. Chapter 5: collecting data. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane handbook for systematic reviews of interventions version 6.5* [Internet]. Cochrane. 2024. Available from: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-05> [Accessed 10 September 2025].
 80. Mathes T, Klaffen P, Pieper D. Frequency of data extraction errors and methods to increase data extraction quality: a methodological review. *BMC Medical Research Methodology*. 2017;17(1):152. doi: [10.1186/s12874-017-0431-4](https://doi.org/10.1186/s12874-017-0431-4)
 81. Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. *Journal of Clinical Epidemiology*. 2006;59(7):697–703. doi: [10.1016/j.jclinepi.2005.11.010](https://doi.org/10.1016/j.jclinepi.2005.11.010)
 82. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. *Journal of Health Services Research & Policy*. 2005;10(1):45–53. doi: [10.1177/135581960501000110](https://doi.org/10.1177/135581960501000110)
 83. Elo S, Kyngäs H. The qualitative content analysis process. *Journal of Advanced Nursing*. 2008;62(1):107–115. doi: [10.1111/j.1365-2648.2007.04569.x](https://doi.org/10.1111/j.1365-2648.2007.04569.x)
 84. Büttner F, Winters M, Delahunt E, Elbers R, Lura CB, Khan KM, et al. Identifying the ‘incredible’! Part 1: assessing the risk of bias in outcomes included in systematic reviews. *British Journal of Sports Medicine*. 2020;54(13):798–800. doi: [10.1136/bjsports-2019-100806](https://doi.org/10.1136/bjsports-2019-100806)
 85. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:I4898. doi: [10.1136/bmj.14898](https://doi.org/10.1136/bmj.14898)
 86. Pluye P, Gagnon M-P, Griffiths F, Johnson-Lafleur J. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *International Journal of Nursing Studies*. 2009;46(4):529–546. doi: [10.1016/j.ijnurstu.2009.01.009](https://doi.org/10.1016/j.ijnurstu.2009.01.009)
 87. JBI. Critical appraisal tools [Internet]. Available from: <https://jbi.global/critical-appraisal-tools> [Accessed 15 September 2025].
 88. Critical Appraisal Skills Programme. CASP checklists [Internet]. Available from: <https://casp-uk.net/casp-tools-checklists/> [Accessed 15 September 2025].
 89. Gates M, Gates A, Duarte G, Cary M, Becker M, Prediger B, et al. Quality and risk of bias appraisals of systematic reviews are inconsistent across reviewers and centers. *Journal of Clinical Epidemiology*. 2020;125:9–15. doi: [10.1016/j.jclinepi.2020.04.026](https://doi.org/10.1016/j.jclinepi.2020.04.026)
 90. Campbell M, Katikireddi SV, Sowden A, Thomson H. Lack of transparency in reporting narrative synthesis of quantitative data: a methodological assessment of systematic reviews. *Journal of Clinical Epidemiology*. 2019;105:1–9. doi: [10.1016/j.jclinepi.2018.08.019](https://doi.org/10.1016/j.jclinepi.2018.08.019)
 91. Ryan R. Cochrane consumers and communication review group. In: Data synthesis and analysis [Internet]. Cochrane

- Consumers and Communication Review Group. 2016. p. 1–7. Available from: https://cccr.org/sites/cccr.org.cochrane.org/files/uploads/AnalysisRestyled_FINAL%20June%202020%202016.pdf [Accessed 20 September 2025].
92. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ*. 2020;368:16890. doi: [10.1136/bmj.l6890](https://doi.org/10.1136/bmj.l6890)
 93. McKenzie JE, Brennan SE. Chapter 12: synthesizing and presenting findings using other methods. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane handbook for systematic reviews of interventions version 6.5* [Internet]. Chichester: Cochrane. 2024. Available from: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-12> [Accessed 15 September 2025].
 94. Becker B.J. Combining significance levels. In: Cooper H, Hedges LV, editors. *The handbook of research synthesis*. New York: Russell Sage Foundation. 1994. pp. 215–230.
 95. Cumpston MS, Brennan SE, Ryan R, McKenzie JE. Synthesis methods other than meta-analysis were commonly used but seldom specified: survey of systematic reviews. *Journal of Clinical Epidemiology*. 2023;156:42–52. doi: [10.1016/j.jclinepi.2023.02.003](https://doi.org/10.1016/j.jclinepi.2023.02.003)
 96. Koricheva J, Gurevitch J. Place of meta-analysis among other methods of research synthesis. In: *Handbook of meta-analysis in ecology and evolution*. Princeton, New Jersey: Princeton University Press. 2013. pp. 3–13.
 97. Schünemann HJ, Vist GE, Higgins JP, Santesso N, Deeks JJ, Glasziou P, et al. Interpreting results and drawing conclusions. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. *Cochrane handbook for systematic reviews of interventions version 6.5* [Internet]. Chichester: Cochrane. 2024. Available from: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-15> [Accessed 20 September 2025].
 98. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the *Journal of Clinical Epidemiology*. *Journal of Clinical Epidemiology*. 2011;64(4):380–382. doi: [10.1016/j.jclinepi.2010.09.011](https://doi.org/10.1016/j.jclinepi.2010.09.011)
 99. Chien PF, Khan KS. Systematic review reporting—writing concisely and precisely. *Pakistan Journal of Medical Sciences*. 2023;39(2):317. doi: [10.12669/pjms.39.2.7428](https://doi.org/10.12669/pjms.39.2.7428)
 100. Aromataris E LC, Porritt K, Pilla B, Jordan Z. JBI manual for evidence synthesis [Internet]. JBI. 2024. Available from: <https://jbi-global-wiki.refined.site/space/MANUAL> [Accessed 20 September 2025].
 101. Clark J, McFarlane C, Cleo G, Ramos CI, Marshall S. The impact of systematic review automation tools on methodological quality and time taken to complete systematic review tasks: case study. *JMIR Medical Education*. 2021;7(2):e24418. doi: [10.2196/24418](https://doi.org/10.2196/24418)
 102. Affengruber L, van der Maten MM, Spiero I, Nussbaumer-Streit B, Mahmić-Kaknjo M, Ellen ME, et al. An exploration of available methods and tools to improve the efficiency of systematic review production: a scoping review. *BMC Medical Research Methodology*. 2024;24(1):210. doi: [10.1186/s12874-024-02320-4](https://doi.org/10.1186/s12874-024-02320-4)
 103. Forbes C, Greenwood H, Carter M, Clark J. Automation of duplicate record detection for systematic reviews: deduplicator. *Systematic Reviews*. 2024;13(1):206. doi: [10.1186/s13643-024-02619-9](https://doi.org/10.1186/s13643-024-02619-9)
 104. Chai KE, Lines RL, Gucciardi DF, Ng L. Research screener: a machine learning tool to semi-automate abstract screening for systematic reviews. *Systematic Reviews*. 2021;10(1):93. doi: [10.1186/s13643-021-01635-3](https://doi.org/10.1186/s13643-021-01635-3)
 105. Motzfeldt Jensen M, Brix Danielsen M, Riis J, Assifuah Kristjansen K, Andersen S, Okubo Y, et al. ChatGPT-4o can serve as the second rater for data extraction in systematic reviews. *PLoS One*. 2025;20(1):e0313401. doi: [10.1371/journal.pone.0313401](https://doi.org/10.1371/journal.pone.0313401)
 106. Scott AM, Forbes C, Clark J, Carter M, Glasziou P, Munn Z. Systematic review automation tools improve efficiency but lack of knowledge impedes their adoption: a survey. *Journal of Clinical Epidemiology*. 2021;138:80–94. doi: [10.1016/j.jclinepi.2021.06.030](https://doi.org/10.1016/j.jclinepi.2021.06.030)
 107. Qureshi R, Shaughnessy D, Gill KA, Robinson KA, Li T, Agai E. Are ChatGPT and large language models ‘the answer’ to bringing us closer to systematic review automation? *Systematic Reviews*. 2023;12(1):72. doi: [10.1186/s13643-023-02243-z](https://doi.org/10.1186/s13643-023-02243-z)
 108. Stone KP, Rutman L, Calhoun AW, Reid J, Maa T, Bajaj K, et al. SQUIRE-SIM (Standards for Quality Improvement Reporting Excellence for SIMulation): publication guidelines for simulation-based quality improvement projects. *Simulation in Healthcare*. 2025;20(2):71–80. doi: [10.1097/SIH.0000000000000819](https://doi.org/10.1097/SIH.0000000000000819)
 109. Cheng A, Kessler D, Mackinnon R, Chang TP, Nadkarni VM, Hunt EA, et al. Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. *Advances in Simulation*. 2016;1(1):25. doi: [10.1186/s41077-016-0025-y](https://doi.org/10.1186/s41077-016-0025-y)
 110. Hopewell S, Chan A-W, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. *The Lancet*. 2025;405(10489):1633–1640. doi: [10.1016/S0140-6736\(25\)00672-5](https://doi.org/10.1016/S0140-6736(25)00672-5)
 111. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet*. 2007;370(9596):1453–1457. doi: [10.1016/S0140-6736\(07\)61602-X](https://doi.org/10.1016/S0140-6736(07)61602-X)
 112. Gruppen LD, Durning SJ. Needles and haystacks: finding funding for medical education research. *Academic Medicine*. 2016;91(4):480–484. doi: [10.1097/ACM.0000000000000983](https://doi.org/10.1097/ACM.0000000000000983)
 113. Kunkler K. Identifying and applying for funding. In: Nestel D, Hui J, Kunkler K, Scerbo MW, Calhoun AW, editors. *Healthcare simulation research: a practical guide*. Switzerland: Springer. 2019. pp. 269–276.
 114. Watts PI, Rossler K, Bowler F, Miller C, Charnetski M, Decker S, et al. Onward and upward: introducing the healthcare simulation standards of best practice. *Clinical Simulation in Nursing*. 2021;58:1–4. doi: [10.1016/j.ecns.2021.08.006](https://doi.org/10.1016/j.ecns.2021.08.006)