Carrying out systematic literature reviews: an introduction

Alan Davies

ABSTRACT

Systematic reviews provide a synthesis of evidence for a specific topic of interest, summarising the results of multiple studies to aid in clinical decisions and resource allocation. They remain among the best forms of evidence, and reduce the bias inherent in other methods. A solid understanding of the systematic review process can be of benefit to nurses that carry out such reviews, and for those who make decisions based on them. An overview of the main steps involved in carrying out a systematic review is presented, including some of the common tools and frameworks utilised in this area. This should provide a good starting point for those that are considering embarking on such work, and to aid readers of such reviews in their understanding of the main review components, in order to appraise the quality of a review that may be used to inform subsequent clinical decision making.

Key words: Health care education ■ Nursing education ■ Nursing research ■ Nursing evaluation research ■ Health care roles

Since their inception in the late 1970s, systematic reviews have gained influence in the health professions (Hanley and Cutts, 2013). Systematic reviews and meta-analyses are considered to be the most credible and authoritative sources of evidence available (Cognetti et al, 2015) and are regarded as the pinnacle of evidence in the various 'hierarchies of evidence'. Reviews published in the Cochrane Library (https://www.cochranelibrary.com) are widely considered to be the 'gold' standard. Since Guyatt et al (1995) presented a users' guide to medical literature for the Evidence-Based Medicine Working Group, various hierarchies of evidence have been proposed. *Figure 1* illustrates an example.

Systematic reviews can be qualitative or quantitative. One of the criticisms levelled at hierarchies such as these is that qualitative research is often positioned towards or even is at the bottom of the pyramid, thus implying that it is of little evidential value. This may be because of traditional issues concerning the quality of some qualitative work, although it is now widely recognised that both quantitative and qualitative research methodologies have a

Alan Davies, Lecturer in Health Data Science, School of Health Sciences, University of Manchester, Manchester, alan.davies-2@manchester.ac.uk

Accepted for publication: June 2019

valuable part to play in answering research questions, which is reflected by the National Institute for Health and Care Excellence (NICE) information concerning methods for developing public health guidance. The NICE (2012) guidance highlights how both qualitative and quantitative study designs can be used to answer different research questions. In a revised version of the hierarchy-of-evidence pyramid, the systematic review is considered as the lens through which the evidence is viewed, rather than being at the top of the pyramid (Murad et al, 2016).

Both quantitative and qualitative research methodologies are sometimes combined in a single review. According to the Cochrane review handbook (Higgins and Green, 2011), regardless of type, reviews should contain certain features, including:

- Clearly stated objectives
- Predefined eligibility criteria for inclusion or exclusion of studies in the review
- A reproducible and clearly stated methodology
- Validity assessment of included studies (eg quality, risk, bias etc).

The main stages of carrying out a systematic review are summarised in *Box 1*.

Formulating the research question

Before undertaking a systemic review, a research question should first be formulated (Bashir and Conlon, 2018). There are a number of tools/frameworks (*Table 1*) to support this process, including the PICO/PICOS, PEO and SPIDER criteria (Bowers et al, 2011). These frameworks are designed to help break down the question into relevant subcomponents and map them to concepts, in order to derive a formalised search criterion (Methley et al, 2014). This stage is essential for finding literature relevant to the question (Jahan et al, 2016).

It is advisable to first check that the review you plan to carry out has not already been undertaken. You can optionally register your review with an international register of prospective reviews called PROSPERO, although this is not essential for publication. This is done to help you and others to locate work and see what reviews have already been carried out in the same area. It also prevents needless duplication and instead encourages building on existing work (Bashir and Conlon, 2018).

A study (Methley et al, 2014) that compared PICO, PICOS and SPIDER in relation to sensitivity and specificity recommended that the PICO tool be used for a comprehensive search and the PICOS tool when time/resources are limited.

Box 1. Main stages of a systematic review

- Define the aims of the study (a set of research questions or objectives)
- Refine and formalise the research question (remove ambiguity and focus on key concepts)
- Break the question down into specific concepts and build a search criteria based on these concepts and related synonyms
- Search literature sources (databases, grey literature, theses)
- Screen literature based on relevance of the title and abstract to the research question
- Screen the resulting full-text literature based on relevance to the research question
- Data extraction (extract details of the studies/quality appraisal)
- Synthesise/analyse the literature
- Report findings

The use of the SPIDER tool was not recommended due to the risk of missing relevant papers. It was, however, found to increase specificity.

These tools/frameworks can help those carrying out reviews to structure research questions and define key concepts in order to efficiently identify relevant literature and summarise the main objective of the review (Jahan et al, 2016). A possible research question could be: Is paracetamol of benefit to people who have just had an operation? The following examples highlight how using a framework may help to refine the question:

- What form of paracetamol? (eg, oral/intravenous/suppository)
- Is the dosage important?
- What is the patient population? (eg, children, adults, Europeans)
- What type of operation? (eg, tonsillectomy, appendectomy)
- What does benefit mean? (eg, reduce post–operative pyrexia, analgesia).

An example of a more refined research question could be: Is oral paracetamol effective in reducing pain following cardiac surgery for adult patients? A number of concepts for each element will need to be specified. There will also be a number of synonyms for these concepts (*Table 2*).

Table 2 shows an example of concepts used to define a search strategy using the PICO statement. It is easy to see even with this dummy example that there are many concepts that require mapping and much thought required to capture 'good' search criteria. Consideration should be given to the various terms to describe the heart, such as cardiac, cardiothoracic, myocardial, myocardium, etc, and the different names used for drugs, such as the equivalent name used for paracetamol in other countries and regions, as well as the various brand names. Defining good search criteria is an important skill that requires a lot of practice. A high-quality review gives details of the search criteria that enables the reader to understand how the authors came up with the criteria. A specific, well-defined search criterion also aids in the reproducibility of a review.

Search criteria

Before the search for papers and other documents can begin it is important to explicitly define the eligibility criteria to
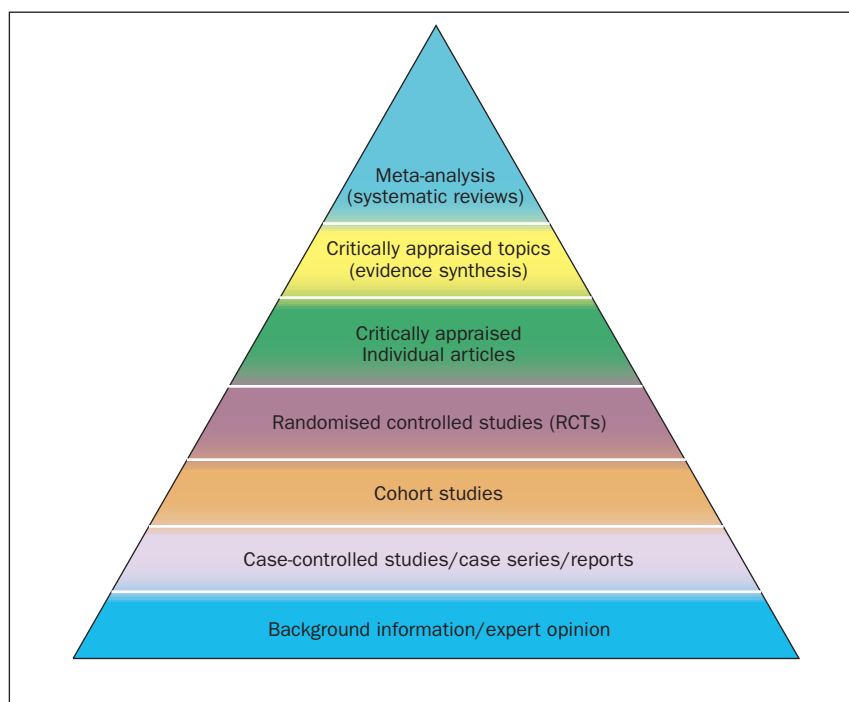

Figure 1. A typical hierarchy of evidence pyramid showing systematic reviews at the top of the evidence hierarchy (adapted from Blunt, 2016)

Table 1. Some common search concept tools/frameworks for developing research questions for systematic reviews

Framework	Components	Primary usage
PICOS	**P**opulation/problem/phenomenon, **I**ntervention, **C**omparison, **O**utcome, **S**tudy design	Used often for medical/health evidence-based reviews comparing interventions on a population
PEO	**P**opulation, **E**xposure, **O**utcome	Useful for qualitative research questions
SPIDER	**S**ample, **P**henomenon of **I**nterest, **D**esign, **E**valuation, **R**esearch type	Often used for qualitative and mixed-methods research questions
ECLIPSE	**E**xpectation, **C**lient group, **L**ocation, **I**mpact, **P**rofessionals, **S**ervice	Policy or service evaluation
SPICE	**S**etting, **P**erspective, **I**ntervention, **C**omparison, **E**valuation	Service, project or intervention evaluation

Table 2. A dummy example using the PICO statement to help formulate a search strategy

PICO element	Concept(s)
Population	- Adult patients - Cardiac - Surgery
Intervention	- Paracetamol - Tablets
Comparison	- Current standard treatment
Outcome	- Pain reduction

Table 3. Some examples of sources of literature often used by nurses and allied health professionals

| Source	Description
PubMed	Life sciences and biomedical topics
Medline	Life sciences and biomedical information
Embase	Biomedical information
Web of Science	Multidiscipline science
Biosis	Life sciences and biomedical topics
PsycINFO	Behaviour and mental health
SCOPUS	Life sciences, social sciences, physical sciences and health science
CINAHL	Cumulative Index to Nursing and Allied Health Literature
Cochrane Library	Database of systematic reviews
CENTRAL	The Cochrane Central Register of Controlled Trials
OpenGrey	Grey literature (conference proceedings, unpublished work)
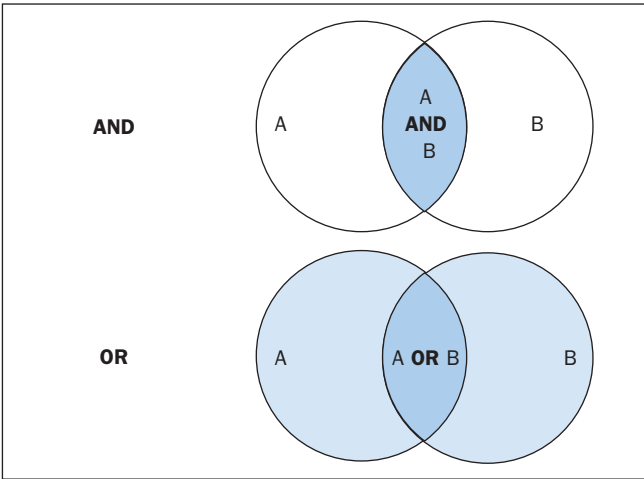
Figure 2. Venn diagrams illustrating Boolean AND, OR groups (blue shading shows results that are included)

Table 4. Common wildcards and operators used in database searches

Wildcard/ operator	Meaning	Example
", { }	Several words	'treatment strategy' {treatment strategy}
#, ?	Alternative spellings or missing characters ie, 'z' or 's' or '-'	visulai#ation visulai?ation
*, $	Truncation, i.e., could include graphs, graphics, graphene etc	Graph* Graph$
AND	Must include both terms	Heads AND toes
OR	Must include one of the terms	Heads OR toes
NOT	Must not have that term included	Graph* NOT photograph

determine whether a source is relevant to the review (Hanley and Cutts, 2013). There are a number of database sources that are searched for medical/health literature including those shown in *Table 3*.

The various databases can be searched using common Boolean operators to combine or exclude search terms (ie AND, OR, NOT) (*Figure 2*).

Although most literature databases use similar operators, it is necessary to view the individual database guides, because there are key differences between some of them. *Table 4* details some of the common operators and wildcards used in the databases for searching. When developing a search criteria, it is a good idea to check concepts against synonyms, as well as abbreviations, acronyms and plural and singular variations (Cognetti et al, 2015). Reading some key papers in the area and paying attention to the key words they use and other terms used in the abstract, and looking through the reference lists/bibliographies of papers, can also help to ensure that you incorporate relevant terms. Medical Subject Headings (MeSH) that are used by the National Library of Medicine (NLM) (https://www.nlm.nih.gov/mesh/meshhome.html) to provide hierarchical biomedical index terms for NLM databases (Medline and PubMed) should also be explored and included in relevant search strategies.

Searching the 'grey literature' is also an important factor in reducing publication bias. It is often the case that only studies with positive results and statistical significance are published. This creates a certain bias inherent in the published literature. This bias can, to some degree, be mitigated by the inclusion of results from the so-called grey literature, including unpublished work, abstracts, conference proceedings and PhD theses (Higgins and Green, 2011; Bettany-Saltikov, 2012; Cognetti et al, 2015). Biases in a systematic review can lead to overestimating or underestimating the results (Jahan et al, 2016).

An example search strategy from a published review looking at web use for the appraisal of physical health conditions can be seen in *Box 2*. High-quality reviews usually detail which databases were searched and the number of items retrieved from each.

A balance between high recall and high precision is often required in order to produce the best results. An oversensitive search, or one prone to including too much noise, can mean missing important studies or producing too many search results (Cognetti et al, 2015). Following a search, the exported citations can be added to citation management software (such as Mendeley or Endnote) and duplicates removed.

Title and abstract screening

Initial screening begins with the title and abstracts of articles being read and included or excluded from the review based on their relevance. This is usually carried out by at least two researchers to reduce bias (Bashir and Conlon, 2018). After screening any discrepancies in agreement should be resolved by discussion, or by an additional researcher casting the deciding vote (Bashir and Conlon, 2018). Statistics for inter-rater reliability exist and can be reported, such as percentage of agreement or Cohen's kappa (*Box 3*) for two reviewers and Fleiss' kappa for more than two reviewers. Agreement can

Box 2. Example search strategy on PubMed

((web OR Internet OR "search engine" OR google OR online OR on line") AND ("help seeking" OR "help-seeking" OR "information seeking" OR "information-seeking") AND (symptom OR symptoms OR diagnoses OR diagnosis))

Mueller et al, 2017

Box 3. Cohen's kappa judgement

- No agreement (<0)
- Slight agreement (0–0.2)
- Fair agreement (0.2–0.4)
- Moderate agreement (0.4–0.6)
- Substantial agreement (0.6–0.8)
- Almost perfect agreement (0.8–1.0)

Adapted from Landis and Koch, 1997

Box 4. Quality appraisal frameworks/tools

- The Jadad/Oxford quality scoring scale: quality of methodology for clinical trials
- Critical Appraisal Skills Programme (CASP): checklist for randomised controlled trails appraisal
- McMaster Evidence-Based Practice Research Group: guidelines for quantitative studies
- Cochrane risk of bias tool

depend on the background and knowledge of the researchers and the clarity of the inclusion and exclusion criteria. This highlights the importance of providing clear, well-defined criteria for inclusion that are easy for other researchers to follow.

Full-text review

Following title and abstract screening, the remaining articles/sources are screened in the same way, but this time the full texts are read in their entirety and included or excluded based on their relevance. Reasons for exclusion are usually recorded and reported. Extraction of the specific details of the studies can begin once the final set of papers is determined.

Data extraction

At this stage, the full-text papers are read and compared against the inclusion criteria of the review. Data extraction sheets are forms that are created to extract specific data about a study (12 Jahan et al, 2016) and ensure that data are extracted in a uniform and structured manner. Extraction sheets can differ between quantitative and qualitative reviews. For quantitative reviews they normally include details of the study's population, design, sample size, intervention, comparisons and outcomes (Bettany–Saltikov, 2012; Mueller et al, 2017).

Quality appraisal

The quality of the studies used in the review should also be appraised. Caldwell et al (2005) discussed the need for a health research evaluation framework that could be used to evaluate both qualitative and quantitative work. The framework produced uses features common to both research methodologies, as well as those that differ (Caldwell et al,
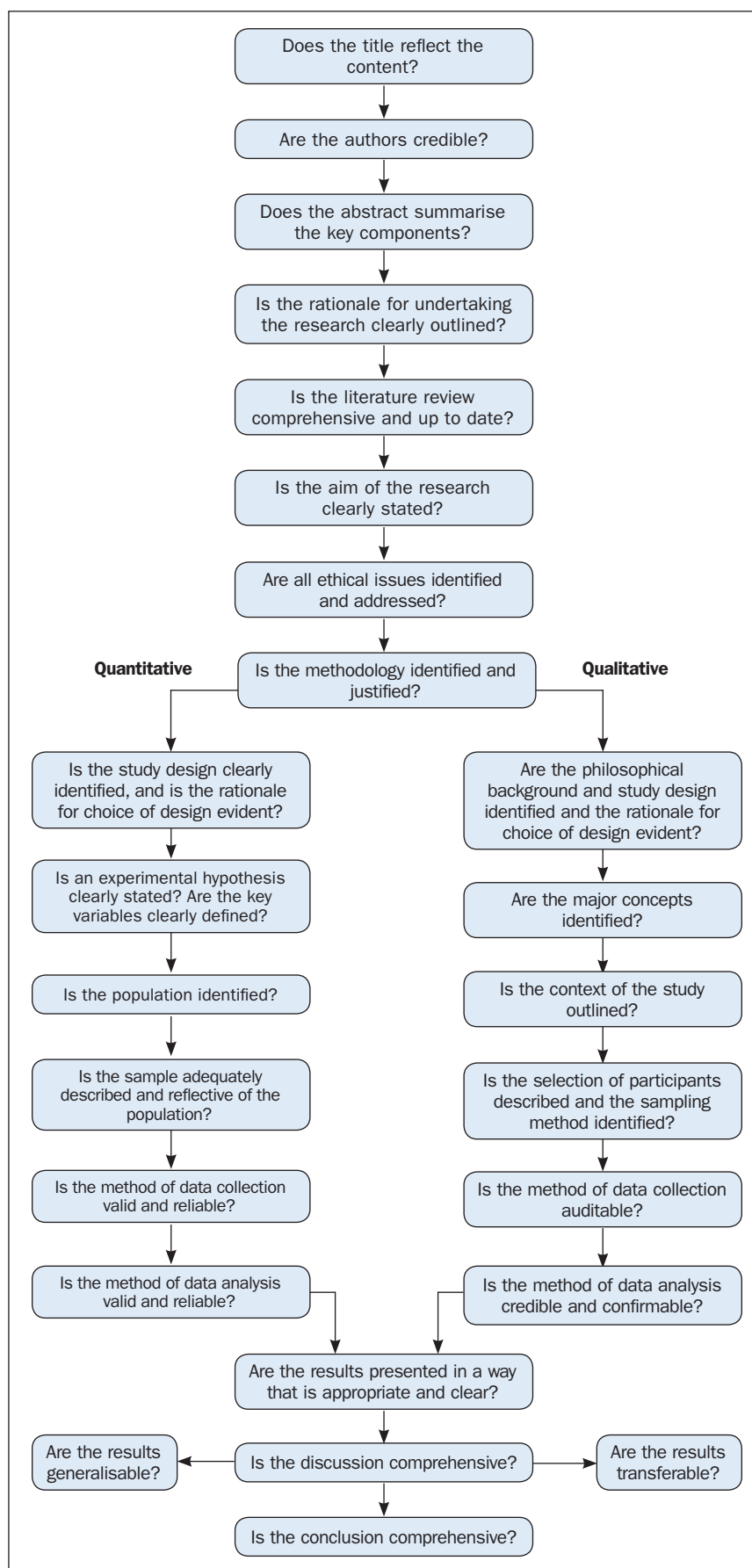
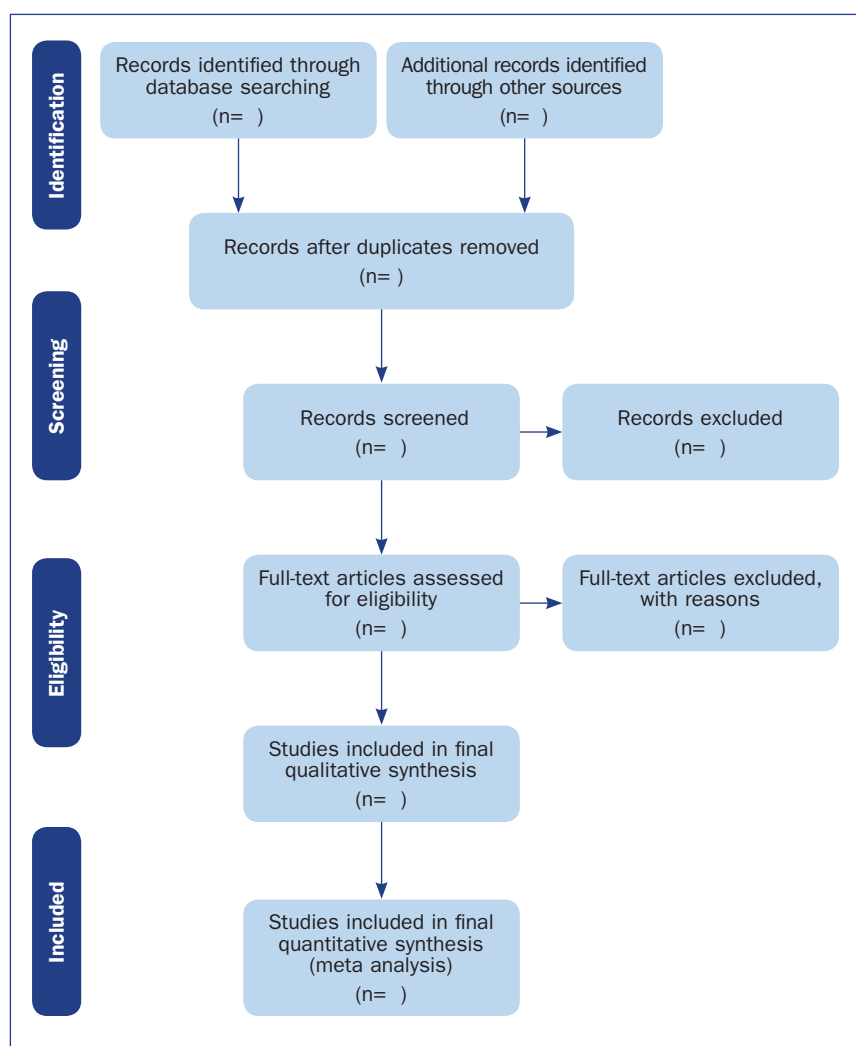Figure 3. Research critique framework (adapted from Caldwell et al, 2005)

Figure 4. The PRISMA flow diagram, where n is the number of items at each stage (Moher et al, 2009)

The PRISMA flow diagram contains the following stages and boxes:

Identification
- Records identified through database searching (n=)
- Additional records identified through other sources (n=)

- Records after duplicates removed (n=)

Screening
- Records screened (n=)
- Records excluded (n=)

Eligibility
- Full-text articles assessed for eligibility (n=)
- Full-text articles excluded, with reasons (n=)

Included
- Studies included in final qualitative synthesis (n=)
- Studies included in final quantitative synthesis (meta analysis) (n=)

James Grellier (wikimedia commons)

	OR
Smith et al, 1991	1.3 (0.5, 2.6)
Jones et al, 1993	2.1 (1.0, 3.4)
Smith et al, 1999	1.8 (0.9, 3.2)
Ng et al, 2004	2.3 (1.9, 2.7)
Chu et al, 2009	2.1 (1.8, 2.5)
Summary measure	**2.2 (1.9, 2.4)**

OR axis: 1.0, 2.0, 3.0

Figure 5. An example of a forest plot showing individual studies on the left and odds ratio on right. The diamond summarises the included studies

2005; Dixon-Woods et al, 2006). *Figure 3* details the research critique framework. Other quality appraisal methods do exist, such as those presented in *Box 4*. Quality appraisal can also be used to weight the evidence from studies. For example, more emphasis can be placed on the results of large randomised controlled trials (RCT) than one with a small sample size. The quality of a review can also be used as a factor for exclusion and can be specified in inclusion/exclusion criteria. Quality appraisal is an important step that needs to be undertaken before conclusions about the body of evidence can be made (Sambunjak and Franic, 2012). It is also important to note that there is a difference between the quality of the research carried out in the studies and the quality of how those studies were reported (Sambunjak and Franic, 2012).

The quality appraisal is different for qualitative and quantitative studies. With quantitative studies this usually focuses on their internal and external validity, such as how well the study has been designed and analysed, and the generalisability of its findings. Qualitative work, on the other hand, is often evaluated in terms of trustworthiness and authenticity, as well as how transferable the findings may be (Bettany-Saltikov, 2012; Bashir and Conlon, 2018; Siddaway et al, 2019).

Reporting a review (the PRISMA statement)

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) provides a reporting structure for systematic reviews/meta-analysis, and consists of a checklist and diagram (*Figure 4*). The stages of identifying potential papers/sources, screening by title and abstract, determining eligibility and final inclusion are detailed with the number of articles included/excluded at each stage. PRISMA diagrams are often included in systematic reviews to detail the number of papers included at each of the four main stages (identification, screening, eligibility and inclusion) of the review.

Data synthesis

The combined results of the screened studies can be analysed qualitatively by grouping them together under themes and subthemes, often referred to as meta-synthesis or meta-ethnography (Siddaway et al, 2019). Sometimes this is not done and a summary of the literature found is presented instead. When the findings are synthesised, they are usually grouped into themes that were derived by noting commonality among the studies included. Inductive (bottom-up) thematic analysis is frequently used for such purposes and works by identifying themes (essentially repeating patterns) in the data, and can include a set of higher-level and related subthemes (Braun and Clarke, 2012). Thomas and Harden (2008) provide examples of the use of thematic synthesis in systematic reviews, and there is an excellent introduction to thematic analysis by Braun and Clarke (2012).

The results of the review should contain details on the search strategy used (including search terms), the databases searched (and the number of items retrieved), summaries of the studies included and an overall synthesis of the results (Bettany-Saltikov, 2012). Finally, conclusions should be made about the results and the limitations of the studies included
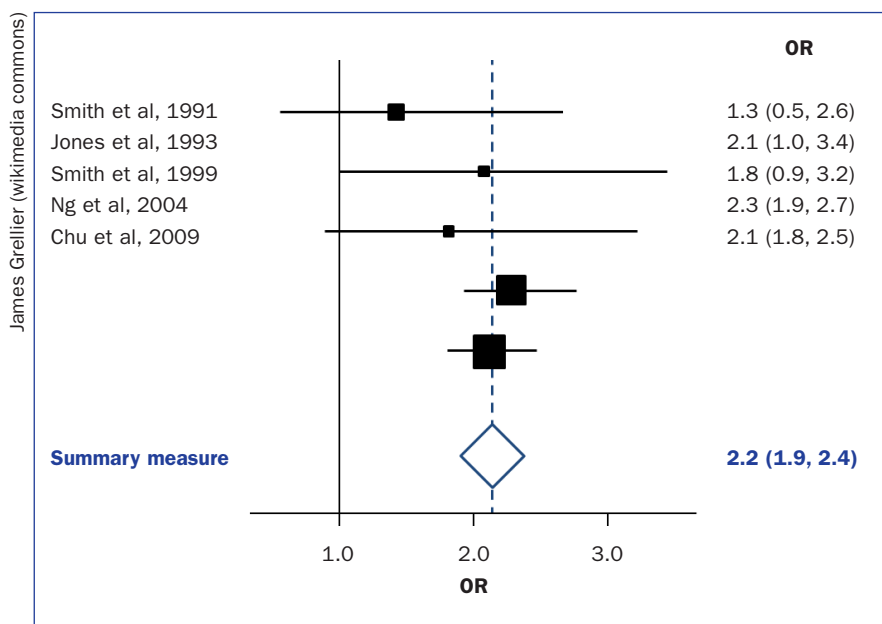
(Jahan et al, 2016). Another method for synthesising data in a systematic review is a meta-analysis.

Limitations of systematic reviews

Apart from the many advantages and benefits to carrying out systematic reviews highlighted throughout this article, there remain a number of disadvantages. These include the fact that not all stages of the review process are followed rigorously or even at all in some cases. This can lead to poor quality reviews that are difficult or impossible to replicate. There also exist some barriers to the use of evidence produced by reviews, including (Wallace et al, 2012):

- Lack of awareness and familiarity with reviews
- Lack of access
- Lack of direct usefulness/applicability.

Meta-analysis

When the methods used and the analysis are similar or the same, such as in some RCTs, the results can be synthesised using a statistical approach called meta-analysis and presented using summary visualisations such as forest plots (or blobbograms) (*Figure 5*). This can be done only if the results can be combined in a meaningful way.

Meta-analysis can be carried out using common statistical and data science software, such as the cross–platform 'R' (https://www.r-project.org), or by using standalone software, such as Review Manager (RevMan) produced by the Cochrane community (https://tinyurl.com/revman-5), which is currently developing a cross–platform version RevMan Web.

Conclusion

Carrying out a systematic review is a time-consuming process, that on average takes between 6 and 18 months and requires skill from those involved. Ideally, several reviewers will work on a review to reduce bias. Experts such as librarians should be consulted and included where possible in review teams to leverage their expertise.

Systematic reviews should present the state of the art (most recent/up-to-date developments) concerning a specific topic and aim to be systematic and reproducible. Reproducibility is aided by transparent reporting of the various stages of a review using reporting frameworks such as PRISMA for standardisation. A high-quality review should present a summary of a specific topic to a high standard upon which other professionals can base subsequent care decisions that increase the quality of evidence-based clinical practice. **BJN**

KEY POINTS

- Systematic reviews remain one of the most trusted sources of high-quality information from which to make clinical decisions

- Understanding the components of a review will help practitioners to better assess their quality

- Many formal frameworks exist to help structure and report reviews, the use of which is recommended for reproducibility

- Experts such as librarians can be included in the review team to help with the review process and improve its quality

Bashir Y, Conlon KC. Step by step guide to do a systematic review and meta-analysis for medical professionals. Ir J Med Sci 2018; 187(2):447–452. https://doi.org/10.1007/s11845-017-1663-3

Bettany-Saltikov J. How to do a systematic literature review in nursing: a step-by-step guide. Maidenhead: Open University Press; 2012:173

Bowers D, House A, Owens D. Getting started in health research. Oxford: Wiley-Blackwell; 2011.

Blunt C. Hierarchies of evidence. 2016. http://cjblunt.com/hierarchies-evidence (accessed 23 July 2019)

Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2008; 3(2):37–41. https://doi.org/10.1191/1478088706qp063oa

Caldwell K, Henshaw L, Taylor G. Developing a framework for critiquing health research. Journal of Health, Social and Environmental Issues 2005; 6(1): 45–54. https://tinyurl.com/y3nulqms (accessed 22 July 2019)

Cognetti G, Grossi L, Lucon A, Solimini R. Information retrieval for the Cochrane systematic reviews: the case of breast cancer surgery. Ann Ist Super Sanita. 2015;51(1):34-39. https://doi.org/10.4415/ANN_15_01_07

Dixon-Woods M, Cavers D, Agarwal S et al. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. BMC Med Res Methodol. 2006;6(1):35. https://doi.org/10.1186/1471-2288-6-35

Guyatt GH, Sackett DL, Sinclair JC et al; for the Evidence-Based Medicine Working Group. Users' guides to the medical literature IX. A method for grading health care recommendations. JAMA. 1995;274(22):1800–1804. https://doi.org/10.1001/jama.1995.03530220066035

Hanley T, Cutts LA. What is a systematic review? Counselling Psychology Review 2013;28(4):3–6

Higgins J, Green S. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. 2011. https://handbook-5-1.cochrane.org (accessed 23 July 2019)

Jahan N, Naveed S, Zeshan M, Tahir MA. How to conduct a systematic review: a narrative literature review. Cureus. 2016;8(11):e864. https://doi.org/10.7759/cureus.864

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1997;33(1):159–174

CPD reflective questions

- Where should high-quality qualitative research sit regarding the hierarchies of evidence?

- What background and expertise should those conducting a systematic review have, and who should ideally be included in the team?

- Consider to what extent inter-rater agreement is important in the screening process

Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res. 2014;14(1):579. https://doi.org/10.1186/s12913-014-0579-0

Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097

Mueller J, Jay C, Harper S, Davies A, Vega J, Todd C. Web use for symptom appraisal of physical health conditions: a systematic review. J Med Internet Res. 2017;19(6):e202. https://doi.org/10.2196/jmir.6755

Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med. 2016;21(4):125–127. https://doi.org/10.1136/ebmed-2016-110401

National Institute for Health and Care Excellence. Methods for the development of NICE public health guidance (third edition). 2012. nice.org.uk/process/pmg4 (accessed 22 July 2019)

Sambunjak D, Franic M. Steps in the undertaking of a systematic review in orthopaedic surgery. Int Orthop. 2012;36(3):477–484. https://doi.org/10.1007/s00264-011-1460-y

Siddaway AP, Wood AM, Hedges LV. How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. Annu Rev Psychol. 2019;70:747-770. https://doi.org/ 0.1146/annurev-psych-010418-102803

Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol. 2008;8(1):45. https://doi.org/10.1186/1471-2288-8-45

Wallace J, Nwosu B, Clarke M. Barriers to the uptake of evidence from systematic reviews and meta-analyses: a systematic review of decision makers' perceptions. BMJ Open. 2012;2(5):e001220. http://dx.doi.org/10.1136/bmjopen-2012-001220