

A Personal Experience of Performing a Systematic Literature Review

M. Bonfield, MSc
L. Fearnside, MSc
F. Cramp, PhD

This paper describes the personal experience of two novice reviewers in developing and conducting a systematic literature review (SLR) in the field of medical ultrasound. The review was performed as part of a PhD study by the first author, which required the researcher to critically review all relevant studies in their specific research area. The review was conducted using a formal, scientific process and highlighted 20 relevant papers. Among the lessons learnt, undertaking an SLR was a time-consuming process. The search produced 1,987 papers, and screening the search yield was facilitated by using a pre-defined protocol. Extraction of data from each study was standardised and generated a 1-page summary of the key findings and methodology to facilitate comparisons of the included studies. The Critical Appraisal Skills Programme (CASP) tool was used for quality assessment. In the critical appraisal process, it is essential that reviewers have good baseline knowledge in the research subject, a solid understanding of the fundamentals of a robust research study including potential bias in research. With the great amount of information that a critical appraisal produces, good information management was key throughout the review to ensure the process was explicit and replicable and so that reviewers could justify their decisions. The methods and processes detailed in this report, as well as its lessons, are relevant to an SLR for any research subject. *J Allied Health* 2018; 47(3):167-171.

GENERALLY SPEAKING, a literature review can be classified into one of two categories: a traditional literature review/narrative review or a systematic literature

M. Bonfield is in Vascular Science, University Hospitals Bristol, NHS Foundation Trust, Bristol; **L. Fearnside** is in Vascular Testing, North Bristol NHS Trust, Bristol; and **F. Cramp** is in Faculty of Health and Applied Sciences, University of the West of England, Bristol, UK.

This report is independent research arising from a Healthcare Science Doctoral Research Fellowship supported by the National Institute for Health Research and Health Education England which was awarded to Michelle Bonfield (NIHR-HCS-D12-03-02). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health. The authors report no conflicts of interest related to this study.

MC1850—Received Mar 30, 2017; accepted Oct 9, 2017.

Address correspondence to: Michelle Bonfield, Vascular Science, University Hospitals Bristol, A225 Level 2 Queens Building, Bristol Royal Infirmary, Bristol, Avon BS2 8HW, United Kingdom. Tel +44 (0)1173423562. michelle.bonfield@uhbristol.nhs.uk.

© 2018 Association of Schools of Allied Health Professions, Wash., DC.

review (SLR). An SLR is defined as “a review of the evidence on a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant primary research, and to extract and analyse data from the studies that are included in the review.”¹ There is a lack of information in the literature on researchers’ experiences of performing an SLR. The aim of this paper was to give an insight into our methods and techniques in performing an SLR in order to enlighten and help others who are embarking on a similar process. Although we solely used the Critical Appraisal Skills Programme (CASP)² quality assessment tool in this study, we believe that the challenges and experiences that are shared in this paper are relevant to the other critical appraisal tools.

Reviews of research are themselves pieces of research and so need to be undertaken according to a defined method. An SLR provides a means of identifying, critiquing, and synthesising existing research evidence using rigorous methodology based on a peer-reviewed protocol.³ Systematic reviews can help to identify gaps between real and assumed knowledge.⁴ Reviewing the literature systematically gives an understanding of the extent of the research activity in a particular subject area and allows us to identify the strengths and weaknesses of how the research has been constructed, analyse the findings, and identify specific characteristics of research foci in this area.³ Carrying out an SLR can point to where improvements in research methods in the subject area are required to identify research gaps and inform new research in this field.⁵

This report presents the personal experience of two reviewers who performed an SLR. The review was designed to summarise and appraise current published and unpublished evidence of the risk factors associated with the development of post-thrombotic syndrome. Included studies were of patients with an objectively confirmed deep vein thrombosis that used ultrasound to quantify residual thrombus and venous reflux. The methods and processes detailed in this report are, however, relevant to an SLR for any research subject.

In accordance with the revised edition of Governance Arrangements for Research Ethics Committees (GAFREC), National Research Ethics Committee (NRES), approval was not required for the study.

Designing the Initial Search Strategy

For the first stage of the SLR, it was considered important to undertake some initial scoping of the literature in the subject area as a mapping exercise. The conceptual framework of the review was developed to explain the key issues to be studied, the constructs or variables, and the presumed relationships between them.⁶

Firstly, the type of study design for the selected research was determined: in this case, observational cohort studies. The mapping exercise had revealed that the observational cohort study was the most commonly implemented study design in this field of research and was considered the most robust design type appropriate to the research question. In designing the inclusion and exclusion criteria for the papers, consideration was given to whether to include only prospective studies or retrospective studies as well and whether only adult studies or also paediatric studies would be included. Points also considered were how the diagnoses were made for the conditions in question (clinical information alone or confirmed by medical imaging/diagnostic tests) and whether the parameters measured were done so objectively using validated tools and by experienced and appropriately qualified staff.

The literature searches were conducted in English, but language was not used as the basis for exclusion. It was considered important to include all potentially relevant research in all languages to ensure that all potentially relevant research was searched for and included where possible. For this review, searches were not restricted by publication date, but depending on the search yield in the scoping exercise, this would be a consideration for any review. For high volume yields, researchers might wish to restrict publications to those from more recent years to include the most current research findings. It would also be important to consider any significant changes in technologies or protocols when considering date restrictions for publications.

The search was developed with assistance from two specialist librarians at the University of the West of England. Most large teaching hospitals in the UK have librarians with expertise in performing literature searches. Although they may not be subject specialists, their expertise in designing a search strategy is extremely valuable across all subject areas. Although in our case the reviewers had both undertaken numerous database searches previously, the volume of literature required to search through in executing this SLR was surprisingly large. The guidance from the specialist librarians in designing the search strategy was invaluable in making this both a more transparent and manageable task. Their experience in dealing with the multiple linguistic variations in the subject area was essential in performing the searches in such a way as to ensure that they were both inclusive yet did not return thousands of irrelevant papers.

It was necessary to take into account variation in the ways study types were understood and labelled to maximise the sensitivity and specificity of the search. Care was taken to search both the controlled vocabulary and free text fields. Search terms combining key terms relating to the subject area were used, and MeSH (Medical Subheadings categories used in the Medline database) terms were used to take into account synonyms, spelling variation, and word forms of the same word. This was especially important in this subject area as, like many medical subjects, many of the terms have several spelling variations. The search strategy was then tested using the “related items” function to ensure that, where relevant studies were not picked up by the original search, these were subsequently identified and an examination of their index terms and language was undertaken to inform how the search strategy could be amended.

The following biomedical and allied health databases were selected: AMED (Allied and Complementary Medicine), CINAHL Plus (Cumulative Index to Nursing and Allied Health Literature), Cochrane Controlled Trials Register, EMBASE, MEDLINE, and Science Citation Index. These databases were selected as they include those that readers would expect to be searched in an SLR in this field, they lend themselves to searching with MeSH headings to nucleate search concepts, and in combination cover significant depth of both international and national first- and second-tier journals to identify all relevant significant publications. Google Scholar was also used for prospective reference list checking to identify research reports and dissertations as well as some journal papers. Web of Science was used to cite additional journal papers³ that may have been missed by Embase and Medline. Hand searching of the reference lists of 18 key review papers and the indexes of two journals which regularly publish studies on the subject in question was undertaken to supplement the search strategy. Generic search engines were used to identify organisations that may publish relevant research. It was considered important to seek out literature from non-electronic sources including contacting two widely published key experts in the field to obtain information on any current studies not yet published, to determine any gaps in the search yield obtained, and to address the issue of the time lag between publication and the indexing of that publication in a database.

Grey literature, including conference proceedings and abstracts from a 20-year period, were searched to identify any current non-commercially published information. Theses from the UK and Ireland were searched using EThOS (Electronic Theses Online Search) and Index to Theses databases. The search strategy was carefully documented and detailed information kept regarding all of the decisions made about how the search engines were used.³

Screening the Search Yield

Studies identified in the searches were then imported into RefWorks web-based bibliographic management software, and any duplicates were removed. Records were kept of duplicates that were removed to account for all of the references identified throughout the review. Good information management needed to be implemented at this stage to avoid lost, misplaced, or uncategorised data that could lead to delays and wasted effort in the review and have the potential to bias the findings. Each item was assigned a unique identifying number enabling the papers to be “queried” in different ways. All paper articles were assigned the same unique identifier that was used in the inclusion/exclusion stage to identify the item. Full records of the search logs and a record of screening at both the title/abstract level and full report level were kept for each document to trace the progress of each item. Records of all attempts to retrieve full reports and why some could not be retrieved were also kept. This was necessary to ensure the process was explicit and replicable and so that the reviewers could justify their decisions. This will also facilitate the updating of the review as more research findings become available in the future. New studies can be incorporated into the review, and it could then be determined if the new information changes the results in any way.

The retrieval of the full-text documents was tracked along with final decisions about whether a study should be included or excluded. A PRISMA diagram was used to account for all references retrieved in the review.⁷ An electronic alert was set up on Mimas Zetoc Alert Service to identify any new publications containing the search terms that were used in the original searches to identify any studies that were published during the time that the review was being undertaken.

A two-stage screening process was used in which the titles were initially screened for suggested relevance and then abstracts of all potentially relevant studies were subsequently screened to identify those that potentially met the inclusion criteria. Where it was not clear whether a paper should be included, full-text papers were obtained for further screening. The screening process was undertaken by the first reviewer alone due to the practicalities of two reviewers independently reviewing the search yield. With hindsight and the appropriate funding, a specialist SLR software tool such as Covidence could have been used to facilitate the independent screening of the search yield and comparison of results between reviewers.

Results of the Search Yield

To provide some context to the extent of the review, from an initial search yield of 2,987 records, 1,980 titles were screened for relevance once duplicates had been

removed, and 281 abstracts and 147 full-texts were subsequently screened for eligibility. Three papers were not held by the British Library, and all possible locations were searched but they remained unobtainable. Three articles were translated from French to English, and 3 were translated from Spanish to English. If, as in our case, no budget existed for translation fees, then this could be seen as an opportunity to enlist the help of any multi-lingual colleagues. One paper published during the data extraction phase of the review was identified as meeting the inclusion criteria and was subsequently added to the review.

In total 20 papers met the inclusion criteria. Two subject experts were contacted and both responded in writing. One commented that the search strategy employed in this SLR was acceptable, and that the relevant studies had been retrieved. The other responded that he was unable to disclose the most important and innovative information that he had currently, as the data would be submitted for publication in a major scientific journal and would be available from spring the following year. Due to the heterogeneity of the included studies, including different assessment techniques and evaluation methods, and large variation in outcome measures, it was not possible to perform quantitative synthesis of the data.

Data Extraction and Critical Appraisal

All literature was assessed for its relevance and quality, and the transparency of its methods and sources was carefully examined. The quality assessment and data extraction were conducted by two accredited Clinical Vascular Scientists. The two reviewers independently undertook data extraction using a data extraction form that was created to enable the standardised extraction of the data from each of the studies as well as facilitating independent data extraction between the two reviewers. This process generated a 1-page summary of the key elements of each study which facilitated the comparison of not only the findings but also the methodology of the included studies. This summary also served as a useful reminder for each study, which aided discussions between reviewers of the critical appraisal elements. The data extraction method was straightforward, largely due to the clarity of the data extraction form, but time consuming, as it was necessary to include adequate detail to support the critical appraisal and facilitate comparison between studies whilst remaining brief enough to serve as a useful synopsis.

Quality assessment of included studies was undertaken and recorded simultaneously to the data extraction. The Critical Appraisal Skills Programme (CASP)² tool for critical appraisal of cohort studies was selected for this purpose. This tool was selected as both review-

ers were already familiar with the tool and had received training in using the tool previously. The tool has also been evaluated with pilot testing in workshops, including feedback and review of materials, using successively broader audiences.⁸

At first glance, the cohort CASP tool appeared to be a relatively straightforward document, comprising of 3 broad questions which are split into 12 individual questions to help you conclude on the 3 broad questions: “Are the results of the study valid?,” “What are the results?,” and “Will the results help locally?” Each of the screening questions had hint questions to help the user answer “yes,” “can’t tell,” or “no.”

Once using the tool, it quickly became clear that despite the simple pro-forma, the multi-faceted nature of the questions meant that several factors had to be considered to answer each screening question. To give examples: question 2 asked whether the cohort was recruited in an acceptable way. The cohort may have been recruited in an acceptable way, such as consecutive sampling over a defined period; however, the study may have limited detail on inclusion/exclusion criteria which would threaten the study sample. Similarly, sample size may be small and, despite acceptable recruitment, produce a major limitation to the study. Question 9 asked if you believe in the results? You may agree that they authors used appropriate statistical methods and dealt with confounding; however, their study may be fundamentally flawed by their recruitment process. The results may not be fully believed—not due to their statistical methods but more to do with the recruitment and study design.

It would seem that the questions have to be answered “on balance” with a weighing up of the information to decide a yes, no, or can’t tell. Critical appraisal certainly has an element of subjectivity and relies on a thorough and dedicated approach to each question to arrive at an “on balance” answer. This becomes particularly challenging when the critical appraisal involves an independent second reviewer, as ultimately both reviewers have to agree on all 12 questions for each paper. The element of subjectivity increases the chances of the reviewers not arriving at the same conclusion. On completing their individual assessments, both reviewers met on several occasions to compare the findings of their independent critical appraisal for each study, and disagreement was resolved through discussion between the two reviewers. A significant amount of preparation was required for these discussion meetings. Each reviewer would need to re-read some sections of the paper in question at the start of the discussion as a reminder of the content. The data extraction page was useful here to provide a synopsis. The individual reviewers would then each present the justification for their conclusion on each point of disagreement. This was an interesting exercise in itself. On

more than one occasion, the process of verbally justifying the conclusion led to the other reviewer changing their opinion without further discussion. In our case, agreement was reached on all points following discussion, without the need to involve a third party, although provision was made for this should it have been required. After agreement was reached, an analysis was made of the patterns of opinion change to ensure that no bias was present with one reviewer more readily conforming to the opinion of the other. This highlights the importance of both reviewers feeling equally empowered to defend their viewpoint.

It soon became clear in the critical appraisal process that it is essential that both reviewers not only have good baseline knowledge in the research subject, but also a solid understanding of the fundamentals of a robust research study. Question 5 asked if the authors identified all the important confounding factors. The reviewers would need knowledge in the specific field to relay the important confounding factors. A good grasp of recruitment procedures and potential bias in research was also essential and addressed in questions 2, 3, and 4.

Despite both reviewers having a good grasp of the subject area and experience in performing their own research studies, there were still challenges. Questions 7, 8, and 9 all focused on the results of the individual papers. To be able to interpret the paper and decide whether to believe in the results requires an understanding of the studies’ statistical methods. A considerable amount of background reading into unfamiliar statistical methods had to be undertaken in order to comment not only on the suitability of the statistics used but the precision of the results. This took a great deal more time than was envisaged.

Conclusion and Key Recommendations

Neither of the reviewers had previously undertaken an SLR and had not received any formal training in the process. Although both had undertaken previous literature reviews at Master’s level, these had not been performed systematically. Following the completion of the process, there were many useful lessons learnt. Screening the search yield was a lengthy process, but the principle reviewer felt able to apply the inclusion criteria confidently using the pre-defined protocol they had created. Both reviewers felt less confident, however, in interpreting and evaluating the extracted data and felt that they needed to re-study the basic statistical methods commonly used in healthcare research in order to gain a better understanding of the significance of the data in each study.

Possibly the biggest learning point from undertaking the SLR was the time-consuming nature of the process, and this should not be underestimated. This review was

limited to 20 papers, and it was surprising to both reviewers how much time it took to complete. The development of the search strategy and the screening for relevance of the search yield took multiple weeks and was a laborious process that required periods of intense concentration. The critical appraisal of the 20 included studies alone required 240 entries per reviewer in addition to the data extraction process. It is important to be realistic about the amount of time a systematic literature review will take, and in addition to the time required to design and carry out the searches, adequate time should be allocated for both parties to independently review each paper and then come together to discuss and finalise the results. In this case, it took several meetings to complete.

The second point is the sheer amount of information that a critical appraisal produces. Good information management was key throughout the review to ensure the process was explicit and replicable and so that the reviewers could justify their decisions. In order to collate this information effectively, spreadsheets were set up for each reviewer which were colour-coded to highlight entries where the reviewers were not in agreement.

Lastly, choosing a second reviewer who has both a good background in the research field and also an understanding of statistical methods is of paramount importance. It is also critical that, due to the volume of work a systematic review generates, both reviewers feel invested in the project to avoid fatigue-related abandonment and to ensure a good working relationship.

Authors' Contributions. MB designed the review, developed the search strategy, and performed the initial searches. MB and LF independently performed the critical appraisal of the included studies. MB and LF wrote the first draft of the manuscript. FC is lead doctoral supervisor and supported the development of the SLR protocol. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Acknowledgments. The authors thank the team of specialist librarians at the Glenside campus of the University of the West of England for their support in creating the electronic search strategy for the review.

References

1. Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for those Carrying Out or Commissioning Reviews. CRD report no. 4, 2nd ed. NHS Centre for Reviews and Dissemination, Univ. of York; Mar 2001.
2. Public Health Resource Unit. The Critical Skills Appraisal Programme: making sense of evidence. Oxford, UK: Public Health Resource Unit; 2006. Available at: <http://www.casp-uk.net/>.
3. Gough D, Oliver S, Thomas J. *An Introduction to Systematic Reviews*. London: SAGE Publ.; 2012.
4. Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*, ver. 5.1.0. The Cochrane Collaboration; Mar 2011.
5. Petticrew M, Roberts H. *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford, UK: Blackwell Publ.; 2006.
6. Miles MB, Huberman MA. *Qualitative Data Analysis: An Expanded Sourcebook*, 2nd ed. Beverley Hills, CA: SAGE Publ.; 1994.
7. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(6): e1000087. doi: 10.1371/journal.pmed.1000097.
8. National Collaborating Centre for Methods and Tools. Critical appraisal tools to make sense of evidence. Hamilton, ON: McMaster Univ.; 2011, updated 16 Dec 2011. Available at: <http://www.nccmt.ca/registry/view/eng/87.html>.