Conducting a systematic review: Demystification for trainees in health psychology

Mark Forshaw, David Tod, & Martin Eubank

The purpose of this article is to define and detail the steps in conducting systematic reviews for trainees and supervisors. We also offer suggestions garnered from our experiences reading, conducting, publishing, and reviewing such manuscripts. Steps include developing specific questions and inclusion/exclusion criteria, undertaking a multi-strategy literature search, implementing replicable data extraction methods, assessing study quality, and employing transparent procedures for synthesising and presenting results. Suggestions include developing a proposal and having it critiqued, allowing sufficient time to conduct a review, keeping meticulous records, and adhering to established procedures.

Introduction

Systematic reviews involve the application of a transparent and systematic method to synthesise research and reduce reviewer bias (Lichtenstein, Yetley, & Lau, 2008). Transparency emerges from the documentation of decision making procedures and public availability of the data search, extraction, and analysis results. Reviewers attempt to reduce bias by describing and following a predetermined replicable protocol. Systematic review procedures have a history in the medical and health communities to underpin practice guidelines, identify research directions, and evidence consensus statements (Lichtenstein et al., 2008). The employment of systematic reviews in health psychology has been increasingly important since it became a necessary piece of evidence in the demonstration of competences for Stage 2, and equally since the introduction of a specific journal for reviews in health psychology in 2007: *Health Psychology Review*. One reason for their increasing popularity is that systematic reviews are considered of higher quality than traditional literature reviews, a view influenced by the 'black box' characteristic of the latter compared with the transparency of the former. In addition to synthesising knowledge, systematic reviews also allow for a rigorous assessment of the quality of the underpinning research.

Based on experiences garnered from our editorial and educational roles, we have found that there is still some misunderstanding among trainees, supervisors, and educators about what is involved in designing, implementing, writing, and publishing systematic reviews. In this article, we overview the systematic review process and offer suggestions on how to undertake them efficiently. We hope this article will serve to demystify the systematic review and help trainees, and supervisors alike, to conduct them more effectively.

According to the Cochrane Handbook for Systematic Reviews of Interventions, "A systematic review attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made" (Chandler, Higgins, Deeks, Davenport, & Clarke, 2017, p. 5). Chandler et al. further suggest that systematic reviews: (a) have clearly stated objectives and pre-defined criteria for study inclusion and exclusion, (b) have an explicit and reproducible method, (c) involve a systematic search that attempts to identify all eligible studies, (d) include an assessment of study quality and the validity of their findings, and (e) involve a methodical synthesis and presentation of the included studies' characteristics, findings, and quality. Some people consider a systematic review to be the same as a meta-analysis, but they are different. A meta-analysis involves the use of statistical procedures designed to synthesise the results from primary research (Borenstein, Hedges, Higgins, & Rothstein, 2009). It is possible to use these statistical procedures to summarise any set of studies, including those not identified systematically (in which case the results are likely to be inaccurate and biased). It is also possible to conduct a systematic review on qualitative studies (e.g., Paterson, Thorne, Canam, & Jillings, 2001), either on their own or in combination with quantitative investigations.

Steps involved in a systematic review

Chandler et al.'s (2017) description above highlights the steps involved in conducting systematic reviews and includes: (a) developing a set of clear specific questions, (b) using

predetermined inclusion and exclusion criteria for study selection, (c) undertaking a systematic multipronged search, (d) implementing an explicit and replicable method for data extraction, (e) assessing study quality, and (f) employing procedures for synthesising and presenting results. The description implies that these projects are conducted in a linear fashion from start to finish. In practice, most reviews are conducted over multiple iterations of searching, analysis, and synthesis, such as in a realist synthesis (Pawson, Greenhalgh, Harvey, & Walshe, 2004). The process is flexible and investigators may adjust the method to help them reach meaningful answers to their review questions. Similar to primary research, questions drive methods. In the following paragraphs, we will briefly describe these steps, based on our experiences. Our intention is not to document definitively how to conduct a systematic review, but instead share reflections to help trainees and supervisors as they fulfil the requirements of their doctoral level training.

1. Develop clear and specific purposes

We have observed that students sometimes make the purpose or scope of their reviews too vague. Focusing on a specific purpose helps individuals identify what they are most interested in examining, and in developing feasible reviews. For example, rather than 'review cancer coping literature,' a more focused purpose might be to 'synthesise the literature examining the influence of coping interventions in older breast cancer patients.' Focused purposes help professionals make decisions in relation to review development and implementation. For example, based on the above example, the inclusion and exclusion criteria are likely to include specifications of age, interventions only targeted at improving coping mechanisms, and so on.

Well-conducted systematic reviews are time and labour intensive, and they involve an opportunity cost (e.g., an academic might have otherwise spent the time developing teaching materials or conducting another type of investigation). The resource expense involved often needs justification, especially if there is a desire to publish it. Two guiding considerations include: knowledge advancement and real world impact. For example, authors might argue that synthesising the breast cancer coping intervention research will lead to the ability of practitioners to disregard

weaker interventions and focus on those which really seem to work. Further, such information may inform practice guidelines, signifying potential impact.

There are existing tools to help authors frame (and then answer) their review questions. One widespread example is the PICO acronym, standing for Problem/Population, Intervention, Control, and Outcome (Chandler et al., 2017). To illustrate, authors might decide to review research focused on older breast cancer patients (population) using specified coping methods (intervention) in research where its efficacy was compared against acceptable control conditions (control) for its influence on objectively measured health indices (outcome). The PICO acronym is suitable for intervention research and can extend to include qualitative work if we include Study type (i.e., PICOS), while other tools are available for difference types of literature and purposes. For example, SPIDER (Cooke et al., 2012) is effective for qualitative reviews. The acronym stands for Sample, Phenomenon of Interest, Design, Evaluation, and Research Type. In addition to helping authors focus their review questions, these tools can also help in other ways, such as developing inclusion/exclusion criteria and identifying search terms.

2. Identify explicit inclusion and exclusion criteria

The development of explicit inclusion and exclusion criteria is possibly one of the more neglected components in the review process. Similar to the purpose statement, however, inclusion/exclusion criteria may influence the scope and quality of the results (Card, 2011). If not considered, authors may apply implicit criteria without realising or discussing notable consequences. For example, a large proportion of systematic reviews in psychology are limited to articles published in English. Inclusion of articles written in other languages is not normally feasible, because reviewers do not have the ability to read them and translation costs can be prohibitive. Studies published in other languages, however, may be relevant (e.g., Tod & Edwards, 2015). Failure to include these studies may be a form of publication bias. Rather than ignoring limitations, such as language, openly discussing the boundaries of a systematic review assists others in identifying solutions that stimulate future knowledge advancements.

The PICO/S tools described above, along with alternatives, can help authors develop suitable inclusion/exclusion criteria. Additional issues to consider include the articles' publication year, their participants' country or cultural origins, and their design features (Meline, 2006). Establishing clear criteria allows reviews to be replicable, a desirable feature of the genre. Even if replication is not desired (or possible), clear inclusion/exclusion benchmarks help readers interpret the contribution of the review to knowledge. To illustrate, tools exist to assist reviewers in evaluating the methodological quality of included studies. Some reviewers might decide to exclude studies based on the methodological assessment due to their desire to summarise only the highest quality evidence available. Other writers argue that including all studies that have been located, and then assessing the influence that methodological features have had on results, may inform interpretations or future research design (Card, 2011). The previous example illustrates that developing inclusion and exclusion criteria is not straightforward. Nevertheless, when combined with a focused review question, time spent developing explicit criteria can assist the next step: the literature search. Given the transparency expected in these types of articles, journal reviewers and Stage 2 assessors alike would expect to see justification regarding the stated inclusion/exclusion criteria.

3. Undertake a systematic and multi-strategy search

It is necessary to use electronic databases to identify relevant studies due to the sheer amount of literature that exists today. Electronic searching is a skill that develops with practice, and some review teams consult or include librarians, as these specialists make valuable contributions to projects. Investigators typically search multiple databases, because such archives do not fully overlap with each other due to variation in which sources are searched. Examples of common databases used in health psychology include Web of Science, PubMed, PsychINFO, and PsychARTICLES. It is also advisable to include databases that reference 'grey' literature, such as Open Grey and PsycEXTRA, and possibly examine research published in PhD theses via the database E-ThOS.

It is beyond our scope to provide guidance on database searching, because there are several components and each involves decision making and piloting. For example, identifying keywords may require discussions with stakeholders as well as some pre-testing. In addition, search strategies often need to be tailored to specific search engines (Egan, MacLean, Sweeting, & Hunt, 2012). Databases are not exhaustive, with estimates of the percentage of studies included in systematic reviews coming from them ranging from 40 to 80 per cent (Hopewell, Clarke, Lefebvre, & Scherer, 2007). Best practice involves a combination of database and various manual search methods. One manual search strategy is a review of journals' tables of contents. Another practice is a backward search where reviewers trawl through located articles' reference lists. Some databases allow investigators to conduct a forward search, where articles citing a located study can be viewed. It is also possible to contact recognised authors requesting any unpublished research that might be relevant.

Many (but not all) systematic reviews attempt to locate and synthesise all the relevant literature on a topic. Although a desirable goal, based on our experience, this can be unrealistic for numerous reasons: databases are incomplete, researchers may not disseminate their work publicly, published work (especially older studies) may lack necessary information for them to be included, authors may not respond to emails, and reviewers may not have the funds or ability to include reports written in languages other than their own. The extent to which reviewers have surveyed the total population of relevant studies is unknowable in most cases, but they will enhance the credibility of their work by acknowledging limitations and demonstrating that they have taken the necessary steps to include as much as reasonably possible (Petticrew & Roberts, 2006).

The literature search is not a one-time discrete step. Rather, searching is an ongoing process that occurs throughout the project's lifecycle. Also, authors may find they refine their search strategies as they proceed. They might uncover new keywords, stumble across previously unknown archives, or need to modify their inclusion/exclusion criteria. There is the impression that authors routinely create the optimal search strategy 'off the bat' with the way many systematic reviews are

presented. Perhaps teams of experienced researchers working in fields they know well are able to develop good protocols that work perfectly first time. Trainees and other inexperienced reviewers, however, may need practice and it is likely that their protocols would benefit from pilot testing and modification. It can also build readers' confidence in the rigour of the project if authors implement checks to assess the efficacy of their search, such as comparing results from team members working independently or with previous reviews.

4. Implement an explicit and replicable data extraction method

Data can include located publications' results, citation information, methodological features, conceptual characteristics, and any other attributes necessary to answer the review question. The number of included studies can range from a few to hundreds, depending on the scope of the questions being addressed. Data extraction tools help ensure consistency across time and individuals. Extraction accuracy does not increase with systematic review experience (Horton et al., 2010). Numerous tools are available, ranging from paper-based crib sheets to web-based software packages, but they are only as helpful as the people using them. 'Off the shelf' tools are best considered as starting points, because they often are not entirely suited to the specific project being conducted; pilot work and adjustments may be needed. Issues warranting consideration include initial set up and running costs, versatility, training requirements, portability, and ability to store, analyse, and present data (Elamin et al., 2009).

Students and trainees sometimes underestimate the challenges involved in data extraction. Even with crib sheets and explicit decision-making rules, it may be difficult to obtain data. Frequently, for example, empirical research reports do not contain information that might be expected, such as descriptive statistics. Details may be reported inconsistently, such as different sample sizes being reported in the abstract, methods, and results. When reviewing a large number of studies, there is the challenge of maintaining attention to detail. Having some method of data extraction evaluation can help reviewers check their consistency and accuracy. Stage 2 trainees

could consider, for example, collaborating with other trainees, supervisors, or colleagues in conducting their reviews as a way of checking their progress.

5. Assess study quality and use the resulting information

Most readers will recognise that variation in primary study quality can influence results. Confidence in the method of a study contributes to trust in the findings (Ryan, Hill, Prictor, & McKenzie, 2013). Similarly, the faith readers have in a systematic review will be influenced by the quality of both the included investigations and the synthesis methodology. Numerous critical appraisal tools and checklists exist to help guide investigators. To illustrate, the Cochrane Collaboration has a 'risk of bias' procedure designed to assess internal validity (Higgins, Altman, & Sterne, 2017).

Similar to quality control in primary research, various practices and definitions exist among systematic reviewers. The variation is partly due to the diversity of primary research designs and types of systematic reviews that exist. For example, the Cochrane risk of bias tool is designed for assessing randomised controlled trials, but it is not suitable for qualitative research. For other methodologies, you should consider using either the CASP checklists, which work effectively for reviews of qualitative studies, or the mixed methods appraisal tool (MMAT), which is suitable for synthesising across studies using different methodologies. The range of practices also reflects theoretical perspectives. Some authors, for example, calculate a research quality score, based on the aggregation of a series of items in a checklist. Other reviewers argue that such a score is limited in value and can prevent authors from identifying design features that might be influencing results across a body of literature (Petticrew & Roberts, 2006). Stage 2 candidates need to be able to defend the choices they make regarding critical appraisal.

One common omission from many systematic reviews is the analysis of study quality results. Having assessed study quality, authors vary considerably in the degree to which they make use of this information. Some might list the information when presenting each the characteristics of each study and leave the interpretation to readers, and others may not even present or refer to the

information in the results or discussion sections. In such cases, authors are doing themselves a disservice. They have probably spent considerable time obtaining this information, yet we would wonder why they did so if they then failed to use the data obtained. Their reviews could certainly have made a greater contribution and impact if they had drawn on the information. In a meta-analysis, for example, information from a quality assessment (e.g., experimental design, questionnaire reliability, participant type, control condition) could be used to identify methodological moderators of the observed effect. In the synthesis of qualitative studies, reviewing researchers' adopted methodological theoretical orientations may also help inform the interpretation of results.

6. Explicitly detail the data synthesis and presentation method employed

Another way that we believe authors have done themselves a disservice in the presentation of systematic reviews is by just describing the results of their searches. They might, for example, produce a table that presents the study characteristics, which are then briefly summarised in the text. These articles would have greater impact if they did more than simply assemble the raw results from their searches. A strong systematic review also includes the more creative aspects of synthesising, analysing, and interpreting the raw data. At a minimum, a review might be expected to include: (a) a description of the search results, (b) a presentation of the located studies' characteristics, (c) an attempt to analyse the data, and (d) the authors' interpretation of the analysis. The reader is better able to evaluate the outcomes of the results and discussion if the underpinning decision-making process has been explicitly and transparently stated rather than being left implicit.

Sometimes, reviewers are able to refer to published guidelines on data analysis and presentation. For example, there are several textbooks presenting details on how to conduct a meta-analysis (e.g., Borenstein et al., 2009). Further, given the transparency principle, ideally authors could also replicate methods used in other reviews. One feature of systematic approaches is the scope for creative and new ways of analysing and presenting data. It is permissible, for example, to present results as text, as numbers, and in graphical ways, as long as authors provide a clear,

concise, and informative answer to the question and they conform to requirements of any stakeholder, such as a funding body or journal. Common examples of data analysis in health psychology include meta-analysis, quantitative mapping, and qualitative thematic analysis. Each has advantages and disadvantages and is suited to specific questions and types of literature. For example, although meta-analysis allows for precise estimates of effect size in quantitative research, the procedures do not help reviewers provide rich descriptions about a phenomenon examined in qualitative investigations.

There are several guidelines available to help reviewers write their reports and include the information needed to inform readers and policymakers. One such example is the PRISMA checklist (Preferred Reporting Items for Systematic reviews and Meta-Analyses; Liberati et al., 2009). The checklist contains items that detail the types of information expected to be included in a report, and there are extensions designed for various types of documents, such as protocols and abstracts. Some journals are now asking authors to include the checklist as part of a manuscript submission.

Reflections and suggestions for trainees

Based on our combined experiences of publishing, supervising, and assessing/moderating systematic reviews, we have listed below some suggestions that we think will assist trainees in completing 'passable and publishable' systematic reviews.

Write a protocol and have it critiqued

Good protocols detail, specifically and concretely, the review questions, rationale, methods, analyses, and presentation plans that will result in the final report. Many trainees (and some supervisors) will probably not have previously undertaken a systematic review and will likely be unaware of potential challenges and obstacles. They might also identity topics that have been previously reviewed, might not be worth doing, or may not be feasible. Writing a protocol may help avoid these pitfalls. We have also found that would-be reviewers are typically not specific enough in identifying their questions and methods. Failure to focus these details sufficiently leads to difficulties (sometimes catastrophic) further down the line. A Google search will quickly locate proposal

examples and templates to assist trainees. Once a protocol has been written, trainees can seek formal feedback from knowledgeable peers and supervisors. Even experienced folks may overlook details, especially if reviewing unfamiliar bodies of literature. Constructive feedback will help trainees expand their own knowledge and enhance the likelihood that they will conduct a high quality review.

We have come across a perception that systematic review protocols must be registered with some organisation prior to being undertaken, such as the International Prospective Register of Systematic Reviews (PROSPERO). Although there are benefits to registering systematic reviews prior to being conducted, such as allowing for peer review, we are unaware of any legislation that requires such mandatory action. Trainees could produce a review that satisfied training requirements without registering their protocol. There are situations, however, when funding bodies, journal editors, or other stakeholders might stipulate some form of registration. Trainees hoping to publish their work could avoid difficulties by considering dissemination plans when planning their projects.

Consider the time needed to produce a document of high quality

We have come across colleagues and students who view systematic reviews as 'soft publications' that are easy to do, and a quick way to secure highly valued and cited articles. These attitudes change once individuals embark on the process, realising that it is much harder than they realised. For example, it can be demoralising when an individual suddenly uncovers an obscure database that contains a large number of relevant documents that have not been identified in more conventional catalogues. Good quality systematic reviews are increasingly held in high regard by the research community, so they are worth doing, and worth doing well.

It may be possible to do a systematic review or meta-analysis quickly if the questions and inclusion/exclusion criteria are so narrow that a search yields a small number of studies (we have seen reviews published with an N=4). It is also possible, however, that such reviews may not add a great deal to the literature and the outcomes may not justify the effort expended (which would

include the authors', editors', and reviewers' time). There may not be enough information to provide a meaningful interpretation. There is a trade-off between question specificity and worthiness that individuals need to balance. In situations where there are not many studies, reviewers need to ensure that they have a strong rationale: that the information gained will make a useful increment to knowledge or will have an impact on practice.

Engage in meticulous record keeping

The aim of systematic reviews is to assess the relevant research and provide detailed informative answers to specific questions. Readers expect to see levels of precision not normally associated with traditional literature reviews and to have opportunities to examine and replicate the raw results from data extraction, if desired. Meticulous record keeping is needed to fulfil such expectations. Record keeping might include, for example, the number of hits from a database on a particular date using an identified search strategy, the percentage break down of the reasons studies were excluded, or the number of participants failing to complete intervention regimes. In our experience, the attention to detail needed sometimes dampens trainees' enthusiasm for a project; however, the use of crib sheets during data extraction and spread sheets can scaffold efficiently their time and effort expended. Several websites contain crib sheets that can be downloaded for various types of studies, and software is also available to help trainees keep track of their projects.

Follow a recognised systematic review framework and read existing reviews

Trainees do not need to reinvent the systematic review wheel. Guidelines exist that can provide assistance, and there are many examples of excellent reviews that can provide inspiration. When first learning a skill, individuals benefit from observing models to help them develop a cognitive map of what they wish to produce, and this observation applies to trainees conducting a review. Once individuals have completed several reviews, they may have the confidence to be creative in the way they conduct their reviews in the future. Developing a narrative around a review is best achieved if one has seen it done before, and practised it, like any skill. This takes time, and is an area of professional development that can truly be regarded as continuous.

Conclusion

The purpose of the current article has been to provide an overview of the systematic review process to help trainees, including Stage 2 candidates, complete their qualification. Although there is considerable variety in publicly available systematic reviews that reflects the adaptability of the process to suit different purposes, stakeholder needs, and types of literature, there are also numerous textbooks, websites, and other resources that can help individuals plan, conduct, and write up their projects. We hope that the steps discussed above are informative for trainees, supervisors, and assessors alike in demystifying the systematic review, and alleviate some of the misunderstandings that exist around the systematic review process and outcome so that authors can be satisfied with, and proud of, their end product.

The Authors:

Mark Forshaw, David Tod, Martin Eubank

Faculty of Science, Liverpool John Moores University

References

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to Meta-analysis*. Chichester, England: Wiley.

Card, N. A. (2011). *Applied meta-analysis for social science research.* New York, NY: Guilford.

Chandler, J., Higgins, J.P.T., Deeks, J.J., Davenport, C., & Clarke, M.J. (2017). Introduction. In J.P.T. Higgins, R. Churchill, J. Chandler & M.S. Cumpston (Eds.), Cochrane handbook for systematic reviews of interventions (version 5.2.0): Chochrane. Retrieved from www.training.cochrane.org/handbook.

Cooke, A., Smith, D. & Booth, A. (2012). Beyond PICO: The SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research, 22*(10), 1435-1443. doi: 10.1177/1049732312452938.

Egan, M., MacLean, A., Sweeting, H., & Hunt, K. (2012). Comparing the effectiveness of using generic and specific search terms in electronic databases to identify health outcomes for a systematic review: A prospective comparative study of literature search methods. *BMJ Open, 2*, e001043. doi: 10.1136/bmjopen-2012-001043.

Elamin, M.B., Flynn, D.N., Bassler, D., Briel, M., Alonso-Coello, P., Karanicolas, P. J., . . . Kunz, R. (2009). Choice of data extraction tools for systematic reviews depends on resources and review complexity. *Journal of Clinical Epidemiology, 62*, 506-510. doi: 10.1016/j.jclinepi.2008.10.016.

Higgins, J.P.T., Altman, D.G., & Sterne, J.A.C. (2017). Assessing risk of bias in included studies. In J.P.T. Higgins, R. Churchill, J. Chandler & M.S. Cumpston (Eds.), Cochrane handbook for systematic reviews of interventions (version 5.2.0): Cochrane Collaboration. Retrieved from www.training.cochrane.org/handbook.

Hopewell, S., Clarke, M., Lefebvre, C., & Scherer, R. (2007). Handsearching versus electronic searching to identify reports of randomized trials. *Cochrane Database of Systematic Reviews, 2*, Art. No.: MR000001. doi: 10.1002/14651858.MR000001.pub2.

Horton, J., Vandermeer, B., Hartling, L., Tjosvold, L., Klassen, T.P., & Buscemi, N. (2010). Systematic review data extraction: Cross-sectional study showed that experience did not increase accuracy. *Journal of Clinical Epidemiology, 63*, 289-298. doi: 10.1016/j.jclinepi.2009.04.007.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Ioannidis, J.P.A., … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ, 339*, b2700. doi: 10.1136/bmj.b2700

Lichtenstein, A.H., Yetley, E.A., & Lau, J. (2008). Application of systematic review methodology to the field of nutrition. *The Journal of Nutrition, 138*, 2297-2306. doi: 10.3945/jn.108.097154.

Meline, T. (2006). Selecting studies for systematic review: Inclusion and exclusion criteria. *Contemporary Issues in Communication Science and Disorders, 33*, 21-27.

Paterson, B.L., Thorne, S.E., Canam, C., & Jillings, C. (2001). *Meta-study of qualitative health research: A practical guide to meta-analysis and meta-synthesis*. Thousand Oaks: Sage.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2004). *Realist synthesis: An introduction*. Manchester: University of Manchester.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Blackwell.

Ryan, R., Hill, S., Prictor, M., & McKenzie, J. (2013). Cochrane consumers and communication review group: Study quality guide. Retrieved from http://cccrg.cochrane.org/authorresources

Tod, D., & Edwards, C. (2015). A meta-analysis of the drive for muscularity's relationships with exercise behaviour, disordered eating, supplement consumption, and exercise dependence. *International Review of Sport and Exercise Psychology, 8*, 185-203. doi: 10.1080/1750984X.2015.1052089