Conducting Systematic Reviews to Inform Occupational Therapy Practice

Susan L. Murphy, Jennifer C. Robinson, Susan H. Lin

Systematic reviews (SRs) are increasingly used in the health professions to evaluate research evidence to guide practice and justify reimbursement for services. Despite the importance of SRs in the health professions, there is no definitive guide for how to conduct a useful, high-quality review. In this article we will (1) provide an overview of the process of writing a traditional SR with particular emphasis on design and conduct, (2) discuss limitations of the traditional SR in occupational therapy, and (3) describe how information is synthesized and used for clinical practice.


Steps Involved in Conducting an SR

Choosing and Formulating a Relevant Research Question

Figure 1 shows the steps involved in conducting a traditional SR. One of the most critical aspects of conducting a SR is choosing a relevant research question. Because time and resources are limited, a research question should be evaluated for its importance to clients or clinicians, its potential for the intervention to be used, and its feasibility of assessment (Counsell, 1997). There should be a clear rationale for
Define the question.
- Specify inclusion and exclusion criteria.
- Specify population.
- Specify intervention or exposure.
- State outcome.
- Clearly specify methodology for systematic review.

Conduct a literature search.
- State search terms.
- Specify databases used (e.g., MEDLINE, Cochrane, CINAHL, PsycINFO, EmBase, Web of Science).
- Specify limits (e.g., language, years).
- Conduct manual searches of key journals.
- Search “gray” literature (e.g., unpublished theses and dissertations, abstracts, conference proceedings).
- Conduct reference citation searches from retrieved literature.
- Obtain titles and abstracts.

Appraise the literature.
- Evaluate abstracts.
- Obtain full articles for potential eligible studies based on abstracts.
- Apply inclusion and exclusion criteria.

Data abstraction.
- Use predetermined abstraction form.
- Extract information on sample, setting, interventions, study design, results, study quality.
- Assess agreement between abstractors.

Analysis.
- Explore heterogeneity of study designs.
- Explore heterogeneity of results.
- Assess study biases.
- Pool or synthesize results.
- Determine practice implications.

Figure 1. Steps Involved in Conducting a Traditional Systematic Review


why the study is being conducted with the goal of improving occupational therapy practice.

To help formulate a good research question for a SR, the question is frequently written in four components, often described by the acronym PICO (IOM, 2008): the population or patient group studied, the intervention, comparison or control group, and outcomes to be addressed. For example, a research question integrating these elements may be “Do stroke patients (P) receiving a protocol of constraint-induced movement therapy (I) have better upper-extremity movement and function (O) compared with stroke patients receiving usual care (C)?” The formulation of a research question has been described by Montori, Swiontkowski, and Cook (2003, p. 45) as an “iterative process” in which the scope of the question is often modified based on the feasibility of the review. For example, a research question may become broader in focus if very few studies can be found on the original topic. In addition, a research question may become more refined if a literature search reveals a group of articles on a specific comparison treatment (e.g., a modified constraint protocol) or addition of other outcomes (e.g., patient satisfaction). Whether the question is broad or more focused, specifying the four components helps clarify the question and improve the quality of each remaining step of a SR.

Establishing a Study Protocol

The study protocol outlines the methods for literature searching, study inclusion, data extraction, and analysis. For example, only studies published in English during a specified time period of 10 years might be included in setting up the protocol. The search strategies such as which databases are to be used; whether unpublished materials, reports, or dissertations are to be included; which key terms are to be searched; and how relevant research will be extracted are all necessary components of the protocol to include so that the results are reliable and can be replicated.

A well-planned protocol is important because it affects bias, one of the main threats to SR quality. Different types of biases affect internal validity of studies. One type is selection bias, in which there are differences between the two groups being compared (IOM, 2008). Researchers seek to control selection bias by randomizing participants into groups for comparison. Usually, a table is presented in which the two groups are compared according to pertinent characteristics such as demographics, function, psychosocial factors, and health status. Differences in outcomes in an exercise study of older adults, for example, may be affected by selection bias if participants in the exercise intervention were more motivated to become regular exercisers or were more active than participants in the control condition at baseline. Another source of bias is attrition bias. In this bias, those participants who completed an intervention study may be different from those who have dropped out during the study. Researchers attempt to control for this bias by doing intent-to-treat analyses to account for all people in the study, regardless of whether they completed or did not complete the study.

In addition to these biases, there are those inherent in SRs. A well-defined protocol will help control for these biases. For example, a comprehensive search strategy of published and unpublished literature is needed to address publication bias (the increased likelihood of studies with positive findings being published). It is also important to be aware that SRs can be biased by differences in how results are reported and how they can be extracted for review (i.e., published abstracts). Authors of SRs may be unknowingly biased even if a detailed protocol is laid out. One way to reduce investigator bias is to think through reasons why there
Inclusion and Exclusion Criteria

It is important to set inclusion and exclusion criteria for studies a priori. Although the scope of the criteria depends on the research question, there are some challenges in devising appropriate prescribed procedures. If criteria are too broad, there is an increased risk of finding disparate studies with varying results that may be difficult to synthesize and to reach conclusions applicable to any single population (Wright, Brand, Dunn, & Spindler, 2007). Although the criteria should be narrow enough to assess an important phenomenon of interest, criteria that are too narrow can (1) bias a review toward false-positive or false-negative results (Counsell, 1997), (2) identify too few studies, and (3) have limited generalizability. Exclusion criteria are typically fewer than inclusion criteria and may include timeframe of studies, articles published in different languages if no translation is possible, or unpublished data (Montori et al., 2003).

Search Strategy

In addition to defined inclusion and exclusion criteria, a preplanned search strategy is necessary. Literature searches often include a variety of techniques such as using multiple electronic databases; examining conference proceedings, books, and unpublished theses and dissertations; manual searching of relevant journals; and perusing reference lists of retrieved materials. It is important to be comprehensive and to provide sufficient detail about how searches are conducted for SR reproducibility.

Another common way to control investigator bias during the literature search is to have a team of investigators agree on which studies will be selected for inclusion in the SR. Often two members of the team independently evaluate studies for selection and quality (Meade & Richardson, 1998), and the interrater agreement is assessed and reported on. The total number of articles found, the specific search terms used, databases and other sources included in the search, and reasons for excluding studies should be reported.

In the first phase of the search, abstracts and titles of studies are evaluated. Full articles that appear to meet criteria on the basis of the abstracts are then retrieved and fully evaluated.

Data Abstraction, Reporting Results, and Conclusions

Once the search has been completed and research reports retrieved, the information has to be abstracted and compiled. Data abstraction is the process of pulling out and recording pertinent information for study comparability (Montori et al., 2003) and is recorded on a data abstraction form. The types of information included are description of participants, interventions used (including the setting and timing of the intervention, who conducted treatment, and whether group or individual treatment was used), study design (e.g., randomized controlled trials [RCTs] or quasiexperimental), and results.

Another main aspect of data abstraction forms is an evaluation of methodological quality. Several systems can be used to evaluate evidence for SRs, and most are hierarchical. In traditional evidence hierarchies (such as the one cited by Holm, 2000), SRs of RCTs or meta-analysis are considered the strongest evidence; controlled trials without randomization and well-designed case-control and cohort studies are considered to be moderate evidence; and single descriptive designs, qualitative studies, and expert opinion are considered to be the weakest evidence.

The next phase of the SR is to draw conclusions from the data extracted in the previous step. For SRs, this involves a synthesis of overall results in which studies should be weighted by study quality as well as for elements we discuss later. Conclusions from an SR should be directly supported by the results. In occupational therapy, there should be a discussion of how the results are clinically applicable and to whom they can be applied to help bridge the gap between research and practice. For example, there may be a lack of clinically significant outcomes measured in the synthesized literature or there may be limited studies in a specific practice area or setting.

Considerations for Conducting SRs in Occupational Therapy

In the preceding sections, we outlined a guide for conducting SRs derived primarily from the medical field. Using a traditional hierarchy of evidence, SRs can help answer a clinical question written in a PICO format or even guide policy decisions. Fortunately the Cochrane Collaboration has spearheaded the process and product development of systematic reviews in health care. The Cochrane Library houses systematic reviews, a registry of controlled trials, and a bibliography of systematic reviews and methodological articles (Cook et al., 1997). Health care professionals are increasingly using levels of evidence and grades or recommendations to interpret research studies and SRs. The advantages to these medical SRs include the following: (1) improving understanding of conflicting evidence (studies with different results), (2) yielding a more convincing and powerful conclusion if the SR is a meta-analysis, (3) explaining whether an intervention is more appropriate for certain subgroups of clients, and (4) helping spotlight where research is needed.

Despite the usefulness of SRs in medicine, clear differences between the medical field and occupational therapy...
should be considered when conducting a SR in occupational therapy. Compared with occupational therapy, the medical field has a larger evidence base with many more RCTs and studies of intervention effectiveness; the focus is generally on drug treatments or clinical procedures to cure disease, and the syntheses usually involve single-component interventions with clearly defined outcomes. Because occupational therapy’s focus is to improve function and reduce disability with treatment targeted to the individual’s resources, context, and abilities, often without a total cure of the health problem, interventions are often more complex and target multiple outcomes. Consequently, the evidence that supports occupational therapy practice may also come from qualitative, correlational, or descriptive studies. Given these differences in evidence bases and types of interventions, we highlight two major considerations when conducting an SR in occupational therapy. The first issue pertains to searching for relevant articles to answer a research question given the heterogeneity of occupational therapy interventions. The second issue pertains to evaluating the evidence despite the paucity of RCTs in occupational therapy. In the following sections, we propose suggestions to deal with these issues when conducting an SR.

**Searching for Relevant Articles to Answer a Research Question**

As mentioned earlier in this article, when we discussed formulating a research question, there is often an iterative process that happens in which an original research question is refined on the basis of the available literature to synthesize. However, the dilemma for the person conducting the SR is determining whether there is an appropriate base of literature from which to draw. Major problems that affect literature synthesis in occupational therapy and other related fields include the heterogeneity of interventions and the use of multiple outcome measures (Jackson & Waters, 2005; van den Ende, Steultjens, Bouter, & Dekker, 2006). One way to address heterogeneity of interventions is to have a panel of experts or an advisory board come to a consensus on how to categorize articles, such as determining whether an intervention is within the domain of occupational therapy or not (van den Ende et al., 2006). Another suggestion for reducing heterogeneity is the attempt to group articles by a theory or theoretical framework (Jackson & Waters, 2005). For example, researchers and clinicians who specialize in sensory integration (SI) have identified characteristics of SI therapy and have developed a treatment fidelity form (Parham et al., 2007). Researchers conducting an SR on SI therapy may exclude a study because the treatment described does not correlate with the essential characteristics of SI therapy.

In addition to addressing heterogeneity of interventions, it is also problematic to synthesize intervention studies when multiple outcome measures are used. Given that this is a common issue in occupational therapy research, van den Ende and colleagues (2006) recommended that people undertaking SRs restrict the outcomes of interest and prioritize primary and secondary outcomes. As the occupational therapy evidence base increases, the designers of RCTs may also aid in the synthesis of research by clearly defining primary and secondary outcomes, as suggested by Nelson and Mathiowetz (2004).

**Evaluating Articles With an Expanded Hierarchy of Evidence**

A traditional SR is a quantitative synthesis weighted heavily toward research quality in which the gold standard is an RCT. Although RCTs are numerous in medicine, there are few RCTs in occupational therapy. A hierarchy of evidence provides a way to help interpret study rigor and quality; however, it has been argued that these types of rating systems discount evidence gained through other research methods (Evans & Kowanko, 2000; Jones, 2004). SRs in occupational therapy often include nonrandomized studies. Helfand (2005) argued that nonrandomized studies should be included in SRs and that nonrandomized studies that are well conducted may be more valid than weak RCTs. In any case, studies gathered for a traditional SR need to be thoroughly evaluated for quality, despite design differences, and justification for inclusion of studies should be reported.

Although we think that the traditional SR approach is best used when the research question involves effectiveness of a treatment on an outcome or outcomes, this approach is somewhat unidimensional in representing often complex therapeutic interventions and is entirely dependent on the hierarchy of the levels of evidence. These limitations can hinder the application of research evidence to practice because large bodies of research evidence gained through studies considered less rigorous are frequently excluded.

For example, Pearson (1999), of the Joanna Briggs Institute (JBI), advocated for a pluralistic approach to considering evidence, citing the value of qualitative research. In particular, researchers at the JBI view evidence broadly to acknowledge clinicians’ complex and multifaceted questions (Pearson, Wiechula, Court, & Lockwood, 2004). One tool developed by the JBI underscores other elements in which to evaluate interventions: feasibility, appropriateness, meaningfulness, and effectiveness, or FAME (Pearson et al., 2007). The FAME scale (Table 1) is used in conjunction with measures of study quality to evaluate how to translate clinical interventions into practice and may be particularly useful when evaluating evidence. Specifically, evaluating
interventions on the basis of the FAME domains is one way to be attentive to priorities of stakeholders, that is, the people who will potentially use the SR. We believe that including a broader range of study designs will facilitate the evidence base of occupational therapy and add to the body of knowledge on the effectiveness of interventions, which will ultimately help use by occupational therapists.

Conclusion

SRs are an important tool for occupational therapy and other health professions to synthesize evidence, support reimbursement for services, and provide services based on evidence. Traditional SRs are optimal when investigating intervention effectiveness; however, studies appraised for inclusion in an SR should be evaluated not only for study quality but also for other elements that would influence translation into practice such as the feasibility of conducting the intervention, appropriateness, meaningfulness to clients and families, and ethical considerations.

All syntheses should be systematic and use a preplanned strategy that can be reproduced. Further work to build on and expand the evidence base will require higher quality studies and a concerted effort to train researchers to conduct relevant practice-oriented studies.

Acknowledgments

We thank Angela Lyden, MS, and Debra Strasburg, MS, for providing feedback on drafts of this article. This work was partially supported by Susan L. Murphy’s Mentored Research Scientist Award from the National Center of Medical Rehabilitation Research (K01–HD–045293) and by Jennifer C. Robinson’s National Research Service Award from the National Institute of Nursing Research (T32–NR–07073).

References


Table 1. The Feasibility, Appropriateness, Meaningfulness, and Effectiveness (FAME) Scale

<table>
<thead>
<tr>
<th>Grade of Recommendation</th>
<th>Feasibility</th>
<th>Appropriateness</th>
<th>Meaningfulness</th>
<th>Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Immediately practicable</td>
<td>Ethically acceptable and justifiable</td>
<td>Provides a strong rationale for practice change</td>
<td>Effectiveness established to a degree that merits application</td>
</tr>
<tr>
<td>B</td>
<td>Practicable with limited training or modest additional resources</td>
<td>Ethical acceptance is unclear</td>
<td>Provides a moderate rationale for practice change</td>
<td>Effectiveness established to a degree that suggests application</td>
</tr>
<tr>
<td>C</td>
<td>Practicable with significant additional training or resources</td>
<td>Conflicts to some extent with ethical principles</td>
<td>Provides limited rationale for practice change</td>
<td>Effectiveness established to a degree that warrants consideration of applying the findings</td>
</tr>
<tr>
<td>D</td>
<td>Practicable with extensive additional training or resources</td>
<td>Conflicts considerably with ethical principles</td>
<td>Provides minimal rationale for practice change</td>
<td>Effectiveness established to a limited degree</td>
</tr>
<tr>
<td>E</td>
<td>Impracticable</td>
<td>Ethically unacceptable</td>
<td>There is no rationale for practice change</td>
<td>Effectiveness not established</td>
</tr>
</tbody>
</table>


