

## SPECIAL COMMUNICATION

# Conducting Systematic Evidence Reviews: Core Concepts and Lessons Learned

Pat A. Brown, EdD, Mark K. Harniss, PhD, Katherine G. Schomer, MA, Melanie Feinberg, PhD, Nora K. Cullen, MD, MSc, FRCPC, Kurt L. Johnson, PhD

**ABSTRACT.** Brown PA, Harniss MK, Schomer KG, Feinberg M, Cullen NK, Johnson KL. Conducting systematic evidence reviews: core concepts and lessons learned. *Arch Phys Med Rehabil* 2012;93(8 Suppl 2):S177-84.

A systematic review (SR) is an essential component of evidence-based practice, because it synthesizes information on a particular topic that is necessary to inform health-related decision making. The purpose of this article is to document the process of producing a high-quality SR within the field of rehabilitation in contrast to other fields (eg, pharmaceutical research). We describe the notable methodologic challenges to producing high-quality SRs for rehabilitation researchers. Broadly, we outline how the quality of SRs is evaluated and suggest mechanisms for researchers to address potential pitfalls. Because meaningful SRs can and should be conducted in this field, we provide practical guidance regarding how to conduct such an SR. We outline a series of 8 important steps in the production of an SR: forming a committee, creating a development plan, conducting a literature review, selecting articles for inclusion, extracting data, preparing tables of evidence, facilitating external review and publication, and forming conclusions and recommendations. For each step of the SR process, we provide detailed description about the methodologic decisions involved and recommended strategies that researchers can implement to produce a high-quality SR. Using these preestablished steps and procedures as a guideline will not only help to increase the efficiency of the SR process, but also to improve the quality. The availability of high-quality SRs with plain language summaries promotes access to the best quality information for all involved in decision making.

**Key Words:** Evidence-based practice; Rehabilitation; Research design; Review literature as topic.

© 2012 by the American Congress of Rehabilitation Medicine

**A** SYSTEMATIC REVIEW (SR) is the end result of a process, which begins with a clearly formulated question with respect to clinical practice, research focus, or policy, and uses a systematic method to identify, select, evaluate, and

synthesize the available research on a topic. As discussed in Dijkers et al,<sup>1</sup> in this supplement, SRs are a key component of evidence-based practice (EBP) where the summary of the available evidence may support recommendations for clinical practice. Ideally, an SR generates information that can guide the rehabilitation-related decision making of a clinician, consumer, or policymaker.

There are many types of SRs, including interventional, diagnostic, prognostic, and measurement reviews, and the general steps to complete them are similar. This article focuses primarily on SRs of interventions, the most common type. In an intervention review, researchers attempt to identify practices that are supported by the strongest evidence by summarizing findings from a set of individual research studies. By evaluating findings from many studies, researchers can make statements about the strength of the evidence supporting a specific treatment, procedure, or practice. At a minimum, high quality reviews are complete (ie, have included all the relevant research available), transparent (ie, have explained all the details about how the review was conducted and how decisions were made), and peer-reviewed (formally in the article submission phase, and often before that by a peer review of the protocol).

## CHALLENGES IN CONDUCTING SYSTEMATIC REVIEWS OF THE EVIDENCE IN REHABILITATION RESEARCH

Although there have been many articles written about the conduct of SRs (eg, Liberati et al<sup>2</sup>), none have addressed the unique challenges faced by researchers in the field of rehabilitation. The quality and quantity of the research available in a field affects the process by which the review is conducted. In a research area with many studies (eg, pharmaceutical research), SRs can be conducted with relative ease because inclusion and exclusion criteria can be restrictive so that only the strongest studies are included. However, in the field of rehabilitation research, as in many disability-related fields, conducting well-controlled, large scale, and/or multisite studies is challenging and depending on the review topic, there are often few if any high quality studies to include in reviews.

### List of Abbreviations

AAN	American Academy of Neurology
AMSTAR	Assessment of Multiple Systematic Reviews
GCS	Glasgow Coma Scale
GRADE	Grading of Recommendations Assessment, Development and Evaluation
EBP	evidence-based practice
PEDro	Physiotherapy Evidence Database
RCT	randomized controlled trial
SCI	spinal cord injury
SR	systematic review
TBI	traumatic brain injury

From the Department of Rehabilitation Medicine, University of Washington, Seattle, WA (Brown, Harniss, Schomer, Johnson); School of Information, University of Texas, Austin, TX (Feinberg); and West Park Healthcare Centre, Toronto, ON, Canada (Cullen).

Supported by the National Institute for Disability and Rehabilitation Research through the University of Washington Model Systems Knowledge Translation Center (grant no. H133A060070).

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit on the authors or on any organization with which the authors are associated.

Correspondence to Pat A. Brown, EdD, Dept of Rehabilitation Medicine, School of Medicine, Box 357920, University of Washington, Seattle, WA 98195, e-mail: [pabrown@u.washington.edu](mailto:pabrown@u.washington.edu). Reprints are not available from the author.

0003-9993/12/9308S-00967\$36.00/0  
doi:10.1016/j.apmr.2011.11.038

Table 1: AMSTAR: Assessing Methodologic Quality

AMSTAR Item	Brief Description
1. Was an a priori design provided?	The research question and inclusion criteria should be established before the conduct of the review.
2. Was there duplicate study selection and data extraction?	There should be at least 2 independent data extractors, and a consensus procedure for resolving disagreements should be in place.
3. Was a comprehensive literature search performed?	At least 2 electronic sources should be searched. The report must include years and databases used (eg, Central, EMBASE, and MEDLINE). Keywords and/or MeSH terms should be stated and where feasible the entire search strategy (algorithm) should be provided.
4. Was the status of publication (ie, gray literature) used as an inclusion criterion?	The authors should state that they searched for reports regardless of their publication type.
5. Was a list of studies (included and excluded) provided?	A list of included and excluded studies should be provided.
6. Were the characteristics of the included studies provided?	In an aggregated form, such as a table, data from the original studies should be provided on the participants, interventions, and outcomes.
7. Was the scientific quality of the included studies assessed and documented?	A priori methods of assessment should be provided (eg, for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo-controlled studies, or allocation concealment as inclusion criteria).
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	The results of the methodologic rigor and scientific quality assessments should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.
9. Were the methods used to combine the findings of studies appropriate?	For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity.
10. Was the likelihood of publication bias assessed?	An assessment of publication bias should include a combination of graphical aids (eg, funnel plot, other available tests) and/or statistical tests.
11. Was the conflict of interest stated?	Potential sources of support should be clearly acknowledged in both the SR and the included studies.

Johnston et al<sup>3</sup> identified additional challenges unique to the conduct of research in the field of disability and rehabilitation, including the following: (1) the complex nature of the interaction between the person with disability and the social and built environment results in research that addresses a wide range of factors (eg, skills, cognition, family, environment) in many different ways, (2) the emphasis on including the participation of people with disabilities in all phases of research may result in the use of methodologies that are considered less rigorous, (3) the small sample size, individualization, and customization required of many interventions, (4) the difficulty or impossibility of blinding and placebo control in many rehabilitation intervention trials, (5) ethical concerns with the use of control groups, (6) inadequate levels of funding for large multisite randomized controlled trials (RCTs), and (7) the fact that many issues of concern in the field of disability and rehabilitation are simply not manipulable (legal issues, human rights, economic factors). In contrast, other health science fields (eg, pharmaceutical research) allow for tighter experimental control (eg, control groups, blinding, and randomization) and for precise dosage of the independent variable.

Because of these challenges in conducting research, the evidence base supporting the field of rehabilitation is composed of intervention studies that use a variety of different designs, including, but not limited to, case studies, cross-sectional, cohort, and crossover designs. While in the broader field of medicine, RCTs are considered the criterion standard of evidence, in rehabilitation research few RCTs are conducted. For example, we recently analyzed 118 SRs in traumatic brain injury (TBI), comprising 4797 individual studies, and found only 16% of the studies included in the reviews were RCTs and most of these RCTs were pharmacologic or physical therapy interventions. SRs on employment, general medical health, quality of life, substance abuse, violence/aggression, or pain

had included few or no RCTs. Not only are there few RCTs in general, but there are even fewer large scale and/or multisite RCTs with active controls, further limiting the strength of the conclusions in the SR. Authors attempting to conduct an SR in rehabilitation often find that (1) there is less research in the area than they expected, or (2) there are an adequate number of studies, but of less quality than they expected (eg, small sample size, weak research designs). Thus, determining best practice will often require rehabilitation researchers to consider a wider range of evidence types than do researchers from other fields.<sup>4</sup> Examples of SRs in rehabilitation that have resulted in recommendations for clinical practice are available through the American Congress of Rehabilitation Medicine (<http://www.acrm.org>), the American Speech and Hearing Association's National Center for Evidence-Based Practice in Communication Disorders (<http://www.asha.org>), and the Academy of Neurologic Communication Disorders and Sciences (<http://www.ancds.org>). The latter includes 2 seminal SRs: 1 on executive function and 1 examining behavioral and social interventions that were included in our analysis of SRs in TBI.<sup>5,6</sup>

### QUALITY SYSTEMATIC REVIEWS

For authors considering writing an SR, it may be useful to consider how SRs are evaluated. Although SRs have become a primary source for EBP, it is clear that not all SRs are equal in terms of quality.<sup>7-12</sup> Consequently, the practice of examining the quality of SRs is becoming more common.<sup>10</sup> One example of a commonly used method for rating SRs is the Assessment of Multiple Systematic Reviews (AMSTAR). AMSTAR is an 11-item measurement scale validated to assess the methodologic quality of SRs.<sup>10-12</sup> Table 1 provides a summary of this scale.

By considering the criteria for evaluating SR quality (either with AMSTAR or some other metric) and building into their plan a means for addressing these criteria, authors can increase the likelihood that they will produce a quality SR.

Meaningful SRs can and should be conducted despite challenges. In this article, we provide potential authors with some practical guidance for conducting SRs in the field of rehabilitation. Our guidance is based on both the knowledge translation literature in general as well as our experience over 5 years of conducting and supporting the production of 31 SRs in spinal cord injury (SCI), brain injury, and burn injury through the National Institute for Disability and Rehabilitation Research funded University of Washington Model Systems Knowledge Translation Center. Using an SR methodology with preestablished steps and procedures is the key to conducting a high quality SR. In this article we propose a methodology and a series of steps (table 2) for conducting SRs in rehabilitation research, regardless of the methodologic challenges within the field. These steps will be described in more detail to guide authors in making informed methodologic decisions while conducting an SR.

### FORMING THE SYSTEMATIC REVIEW AUTHOR/ REVIEW COMMITTEE

Forming the SR committee is naturally the first step to conducting an SR. This committee includes the authors of the review, but can also include other members with subject matter expertise or other SR experience that can provide consultation during the SR process. Authors with subject matter expertise on the topic chosen for the SR is critical, but authors with less expertise on the topic but with experience with conducting SRs or with a strong statistical background can also be critical on an

SR committee. However, keep in mind another key component to establishing your committee is the rapport among the possible members. SRs often take a year or more to complete and require working together for long periods of time, so ensure members can work well together. The committee should also establish 1 person to be the leader/coordinator to make sure the SR committee members meet regularly and stay on task. Lastly, there should be no conflict of interests regarding the outcome of the review with any of the potential committee members.

### CREATING A DEVELOPMENT PLAN FOR A SYSTEMATIC REVIEW

A high quality review begins with a well thought out plan. An SR plan should include several elements: a clinical question, a description of the study population, the inclusion and exclusion criteria, the intervention(s) to be studied (for an intervention review), the outcomes to be included, research methodologies to be considered, and the grading system to be used. In addition, from the beginning, SR authors should consider the criteria that will be used to evaluate the quality of the review once it is completed.

### HAVE A CLEAR QUESTION OR GOAL

A high quality SR begins with a clearly defined question. The question should be answerable and of interest to the intended audience: clinicians, researchers, or policymakers. The goal of the SR is to improve the quality of care by synthesizing current literature into a usable format, laying a foundation for effective knowledge transfer, and to improve programs and services. The following criteria should be considered when writing questions for an SR:

- What is the purpose of this review (eg, to answer what question[s])?
- Who will this review help (ie, researchers, clinicians, consumers)? Who will use this information (ie, audience)?
- What will the information that may be gained from this review accomplish? Will it inform decisions about clinical practice, practice guidelines, consumer behavior, or policy (including conclusions that there is a lack of evidence to support recommendations)?
- How does the clinical question guide the type of review conducted? Will the studies reviewed be intervention, prevalence, or measurement studies?
- How will the evidence obtained be translated and communicated to the intended audiences?

Responding to these questions and writing a clear question can be time-consuming and may require quite a bit of negotiation among a team of SR authors, but doing it early will help to avoid confusion and challenges later on. It will also focus the team so that authors are not distracted by other interesting, but not directly related, topics that will arise during the SR process. Although writing clear questions will help avoid topic creep, it does not mean that questions are written in stone. In fact, questions may need to change as more information is accumulated and evaluated about the nature of the research that is available. Keep in mind though that when the main questions change, the development plan should be revisited and changed accordingly.

### DEFINE THE STUDY POPULATION

It is important to be specific in defining the population of interest in the SR. At a minimum, this includes the specific type of disability, age range, and sex. There are dangers, however,

**Table 2: SR Steps**

- 
1. Form the SR committee
  2. Development plan
    - Define clear clinical questions and goals
    - Define study population
    - Define inclusion and exclusion criteria
    - Define key literature search terms and explore the literature
    - Review the grading system for the SR
  3. Literature search
    - Using the key words agreed on in the development plan
    - A literature search is conducted using numerous electronic library databases
  4. Selecting articles for inclusion
    - Review abstracts selected in literature search
    - Use at least 2 independent reviewers
    - Use categorization system for inclusion and exclusion
  5. Data extraction
    - Decide how to extract and store the data (ie, web-based, Excel, worksheets)
    - Use at least 2 independent reviewers to extract data
  6. Tables of evidence
    - Facilitate analysis and grading the evidence
  7. External review and publication
    - External reviewers are provided with the tables of evidence to review for completeness, clarity, and to make recommendations
    - Authors finalize the manuscript and submit for publication
  8. Conclusions and recommendations
    - Recommendations made from evidence discovered in the review process, and submitted for final publication
-

in being either too precise or too imprecise in the early stages of planning. For example, if one were conducting an SR in the field of brain injury rehabilitation, then defining one's study population as acquired brain injury, an umbrella term encompassing a wide spectrum of brain injuries including TBI and nontraumatic etiologies such as strokes and aneurysms, might be too broad. If the clinical question focuses on severe TBIs, then one might choose an operational definition such as a Glasgow Coma Scale (GCS) score of 8 or less.<sup>13</sup> However, 1 challenge in defining a study population this precisely is that many studies may not report a GCS score, making it impossible to know the level of severity reported in the study population. Because it can be hard to know in advance what demographic information will be reported, in the development plan it is important to allow for some flexibility and consider what one would do in this case. For example, if studies do not provide a GCS score, another less rigorous definition may be considered, such as including all studies reporting a TBI sample, rather than specifying the level of severity.

### DEFINE THE STUDY INCLUSION AND EXCLUSION CRITERIA

In addition to defining the target population, other criteria for including studies in an SR are defined in the development plan. These include the interventions to be reviewed, the outcomes to be included, the methodologies to be considered, and time frame of the study. Consideration must also be given to which studies should be excluded from the SR. A well-developed plan will incorporate the elements in table 3.

### DEFINE LITERATURE SEARCH TERMS

In an ideal world, authors would be able to perfectly identify all relevant search terms, conduct a search, locate all relevant

articles, and move forward with the review. In reality, in any domain, determining the search vocabulary for SRs is a complex, specialized, iterative process.<sup>14</sup> Our ongoing work in developing rehabilitation-focused reviews has enabled us to identify several focused strategies to facilitate topic conceptualization and subsequent search vocabulary construction. When developing the SR project development plan, we have found it productive to run preliminary searches on the proposed topic area, summarize our findings, and use this material to focus discussion about the topic in general, as well as to identify potential search terms. This initial search can help review authors refine or expand the preliminary topic as necessary, and it also ensures that any conflicts in terminology, among anyone on the project team, or in the overall subject literature, are identified and reconciled.

Vocabulary differences are most likely to occur when proposed topics concern emerging areas. For example, when preparing for an SR on the topic of SCI and urinary surveillance (ie, methods for monitoring urinary tract health), initial searches showed very few results with the term *surveillance*. On discussion with content experts, we operationalized the concept of surveillance to be a series of screening, tests, and tools used separately or in combination to prevent negative urinary outcomes such as infection or cancer. Thus, surveillance could be represented with a combination of terms used for specific urologic and related tests and for specific urologic and related conditions or symptoms. As another example of this type, a proposed topic was initially described as the level of scar satisfaction for people with burn injuries. The term *scar satisfaction* was not commonly used in the literature; however, the study authors were reluctant to expand the literature search to include the commonly used concept of self-esteem, because this construct was not specific to the acceptance of scarring that they most wanted to investigate. In an attempt to maximize both recall and precision in this case, we developed a complicated term structure, combining the term *burn* with either terms specifically related to physical self-perception or the 2 terms *burn* and *scar* with terms related to general perception and self-esteem.

To facilitate conversations about search terms, it can be helpful to identify several studies published on the topic during the initial search. As they review these studies, SR authors can mine them, not just for potentially relevant citations, but for potentially relevant vocabulary (including the terms used by databases such as PubMed to index the articles, as well as terms used in the article itself). Once identified and aggregated, these candidate terms can then be finalized by the SR committee to be used for the final literature search.

### IDENTIFY A GRADING SYSTEM

A grading system is a way to classify individual research studies based on the degree the findings are subject to bias (systematic errors) and threats to validity. Johnston et al stated that, ". . .evidence grading is based on expert knowledge of the typical validity threats associated with various research designs. . .applied to a particular question and study under review."<sup>15(p293)</sup> Grading the quality of the evidence facilitates the move from research to practice. Once the methodologic quality of each included study has been established and the strength of the evidence as a whole has been determined, the SR authors can make recommendations based on this evidence. Practices with strong evidence reported in the SR may be incorporated into clinical practice guidelines and information to assist consumers in decision making.

**Table 3: Sample Inclusion and Exclusion Criteria**

Inclusion and Exclusion Criteria	Example (Hart and Fann, 2009)
What are the interventions to be reviewed?	Interventions to treat depression in those with TBI including drug, therapy, and nontraditional treatments.
What are the interventions to be excluded?	Interventions where participants were not screened for depression using a valid depression assessment tool.
What are the outcomes to be reviewed?	Where depression is measured as an outcome variable.
What are the outcomes to be excluded?	Anxiety, PTSD, stress when not reported as covariates of depression.
What methodologies should be reviewed?	Experimental and observational research that seeks to investigate interventions that treat depression of those with TBI.
What methodologies should be excluded?	Literature reviews, nonpeer-reviewed literature, or expert opinion; nonintervention research.
What is the time frame of publication of the studies to be reviewed?	Studies from 1980 to current.

Abbreviation: PTSD, posttraumatic stress disorder.

Selecting a grading system that will work best for assessing the evidence in an SR can be difficult due to the sheer number of available systems from which to choose.<sup>16</sup> One approach to capturing the best available evidence is to combine multiple grading systems to assess the quality of the included studies. In the Evidence-Based Review of Rehabilitation of Moderate to Severe Acquired Brain Injuries, empirical studies were assessed using the study quality scoring system of the Physiotherapy Evidence Database (PEDro) for RCTs. PEDro<sup>17</sup> is the rating scale developed by the Centre for Evidence-Based Physiotherapy in Australia ([www.pedro.fhs.usyd.edu.au/FAQs/Scale/scaleitems.htm](http://www.pedro.fhs.usyd.edu.au/FAQs/Scale/scaleitems.htm)). Studies that used a nonexperimental or uncontrolled design (nonrandomized comparative trials, cohort studies, or retrospective trials) cannot be evaluated using the PEDro scale and therefore the Downs and Black scale<sup>18</sup> was applied.

In addition to grading each individual study included in the SR, the evidence as a whole is also graded at the conclusion of the study. For example, in the Spinal Cord Injury Rehabilitation Evidence study, once the methodologic quality of each study was determined, the resulting evidence was assigned a level from 1 to 5.<sup>19</sup> Both Grading of Recommendations Assessment, Development and Evaluation (GRADE), a system used in many reviews in evidence-based medicine, and the classification of evidence developed and used in reviews endorsed by the American Academy of Neurology (AAN) support classifying the evidence in individual studies and assessing the evidence as a whole across multiple studies. Using GRADE, the quality of the evidence across multiple studies is assessed separately for each reported outcome of the intervention. This evidence may be assessed as high, moderate, low, or very low quality. A review panel considering the evidence in the SR would then decide (1) which reported outcomes are critical to the recommendation that they are considering and (2) of these outcomes, what is the lowest quality level. The overall quality of the evidence, for the outcomes of interest, would be rated at that lowest level. If the overall quality of the evidence is low, "further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate" (<http://www.gradeworkinggroup.org/>).

Using AAN,<sup>20</sup> the methodology of each included study is rated class I (prospective, randomized, controlled clinical trial with masked outcome assessment, in a representative population) through class IV (evidence from uncontrolled studies, case series, case reports, or expert opinion). Next, to translate the evidence into recommendations, a level of A, B, or C is assigned to the body of evidence as a whole. Level A (the body of evidence in the SR includes at least 2 class I studies) results in a recommendation that the intervention or therapy should (or should not) be done.

Despite the widely accepted use of these grading systems, there are significant issues in applying these systems in rehabilitation research. In GRADE, AAN, and PEDro, as in most evidence grading systems, RCTs are the criterion standard for evidence. A well-conducted RCT includes blinding, that is, a process that prevents subjects (single-blinded) or subjects and researchers (double-blinded) involved in the trial from knowing whether subjects are in the experimental or control groups.<sup>17-21</sup> As Johnston et al<sup>15</sup> point out, blinding is extremely difficult in rehabilitation. For examples and a comprehensive discussion of this and other issues related to evaluating the evidence in rehabilitation research, see Johnston and Dijkers<sup>22</sup> in this supplement.

## CONDUCTING THE LITERATURE SEARCH

Once a project development plan has been completed and the initial literature screening is done to tease out the final search words, SR authors are ready to move forward and implement a literature search using multiple electronic databases. Using a minimum of at least 2 electronic sources or databases is required by AMSTAR to be considered a quality review. A technique to increase the yield of these searches in the different databases, in which our work similarly confirms the general findings of Sampson et al,<sup>23</sup> is to implement tailored search protocols across multiple databases, taking advantage of the controlled vocabulary and other features of each system. We have been able to gradually observe which databases contribute most significantly to our reviews and to adjust the databases used according to their relative utility in contributing unique, relevant articles. We eliminated ProQuest and Google Scholar from our reviews, because they contributed very few relevant additional studies to our list of included studies and contributed no unique results that were highly targeted to review topics. Google Scholar, in particular, contributed excessively large numbers of irrelevant results, and eliminating it increased workflow efficiency without sacrificing overall recall.

## CLASSIFYING STUDIES BASED ON THE INCLUSION CRITERIA

When SR authors are satisfied that they have conducted a comprehensive search, they will find themselves with a long list of references and abstracts. Before acquiring the actual articles, it will be useful to conduct a preliminary selection by reviewing the abstracts of the studies and applying the inclusion and exclusion criteria identified in the SR plan. Determining a study's inclusion based on the abstract can be difficult, especially with abstracts that are minimally descriptive, so at this point it is beneficial to be broadly inclusive (ie, if there is a doubt about whether it should be included or not, authors should include it, get the full article, and review it more fully).

At this point in the review process, it is also useful to categorize all studies into subcategories based on a relevant criterion. This is recommended because in the early stages of writing an SR, the authors' goal may be to only include experimental studies reporting outcomes related to their questions of interest. However, after reviewing the literature they may not find any studies meeting these criteria, but instead find studies that are observational, or nonexperimental. If the authors decide to then broaden the inclusion criteria and include studies with other methodologies, they would have to go back through their articles and apply the new inclusion and exclusion criteria. A more efficient approach is to develop and apply a categorization system for all studies from the beginning, and then if authors want to broaden their inclusion criteria, they can do so easily.

For example, in our work, we have used an inclusion strategy using 3 categories: (1) is the study experimental versus nonexperimental (eg, observational or other design where there is no experimental manipulation) (we actually have subcategories here as well); (2) are study outcomes of primary or secondary interest to the SR; and (3) is it a qualitative study investigating the primary or secondary outcomes of interest. With this type of categorization system, authors have the opportunity to modify their inclusion criteria if they find that the research base in their given topic area is less substantial than they had assumed. Without a categorization process, an SR team would have to abandon the SR or conduct the abstract review again to find relevant literature for their review. A categorization process then is a pragmatic approach to inclu-

sion/exclusion decisions that may be particularly necessary in a field like rehabilitation research that has such a diverse research base.

Exclusion categories may also be applied during the abstract review. Exclusion criteria may include animal studies, studies not available in English, and those reporting expert opinion. It is important to note that to meet 1 of the AMSTAR criteria for a quality SR, it is important to document the studies that are both included and excluded from the SR. This can be done in the article or in a separate document posted online.

Once the final criteria are established, a minimum of 2 reviewers read the abstracts of the studies located in the SR literature search. Using these criteria of inclusion, and exclusion, reviewers independently categorize the articles. Thus, each abstract is reviewed at least twice to determine inclusion status for the SR. Use of bibliographic software or a spreadsheet is recommended to keep track of the categories assigned to the studies.

### EXTRACTING DATA

After categorization is complete and a final list of included studies is available, reviewers begin to fully review and extract data from the included studies. To extract data, the review team needs to make 3 key decisions: (1) what type of data needs to be extracted, (2) how to define the main data types and concepts, and (3) how they will extract and store the data. Depending on the type of review conducted, the type of data collected varies. As noted in the project development plan, sample size, research design, and study population will be collected across all review types but some review topics may need additional and specific data collected. Deciding these fields a priori to data extraction is essential to conducting a quality SR, and keeps reviewers from having to go back to collect additional data later.

Creating a common language and dictionary for the data fields is also very important. In our experience, authors from different disciplines will use different names to describe the same research designs. For example, an author with a social science background might describe a study as a repeated-measures design, while another author might label it as a crossover design. As a result, we have found it useful to develop a common language flowchart decision tree to ensure that all reviewers are classifying the research design in the same manner. Although our design tree was developed internally, there are other design tree tools<sup>24</sup> that could be used to ensure data extraction is consistent across reviewers.

After deciding and defining the final data fields, the next decision is how the SR data will be collected and stored. Data collection for SRs is usually done via worksheets that are then entered into spreadsheets,<sup>25</sup> but this is often very time consuming and inefficient. According to Elamin et al,<sup>25</sup> web-based data applications are more costly to develop but they are generally rated as being easier to use for project setup, versatility, training, portability, ability to manage data, ability to present data, and the ability to store and retrieve data than other methods of data extraction. Our team has developed a web-based data entry system using a Structured Query Language database to store the data for all its SRs. We have found this web-based system to be more efficient, versatile, and user-friendly than the worksheet method. For SR authors who do not have technical support, there are free and commercial products available that also provide this function (eg, RevMan, DistillerSR). Regardless of how a review team decides to extract the data, the key elements to consider are that the system allows for more than 1 reviewer to review each included study, that it is easy to use,

and that the data can be retrieved in a usable format to facilitate the development of the tables of evidence.

Similar to the inclusion/exclusion process, a minimum of at least 2 reviewers should extract data from each article. Differences in article data between the 2 or more reviewers should be reconciled through consensus.

### CREATING TABLES OF EVIDENCE

Once data are collected and extracted from articles, tables with the main data components are developed. The main purpose of these initial tables is to facilitate analysis and summarization of the findings across the research articles, but a secondary purpose is to facilitate grading of the evidence. Authors and committee members use the data in these tables to finalize the evidence grade. At this stage of compiling the evidence, SR teams will often create 3 main types of tables and reports: (1) descriptive/detailed tables, (2) final evidence tables, and a (3) technical report.

The descriptive/detailed tables are comprehensive and usually report all of the data extracted from the articles. These tables are often not meant for publication, because they are very lengthy and provide too much detail about the articles. However, they are very useful for initial analysis and facilitate grading the evidence.

The final evidence tables are usually high-level summaries of the descriptive and detailed tables and are designed to be included in the final publication. Key features of these tables include sample size, a description of the sample, the research design (including masking and blinding information), a brief description of the intervention(s), the main outcomes, how the outcomes were measured, and the major conclusions of the study. This table should also include the level of evidence assigned by the review team to each article.

Technical reports provide full documentation of the methodology used in the SR. These reports include both the descriptive and final evidence tables. They also provide the detailed development plan for the SR, the inclusion categorization system, and data dictionaries or guides that may have been created for data collection (ie, a research design tree). A final list of included and excluded articles is also usually provided in these very detailed reports.

### EXTERNAL REVIEW AND PUBLICATION

After data extraction and tables of evidence are complete, writing the article for publication is the next step. At this point, a review of the methodology, the project development plan, and the resulting tables of evidence should be reviewed by at least 2 external reviewers, as required by the AAN guidelines.<sup>20</sup> Finding 2 available external reviewers in addition to the SR authors may be difficult and so it is recommended that the SR team nominate 2 to 4 external reviewers to ensure that at least 2 of them can complete the review. Nominated external reviewers are more likely to participate if there is enough information about the SR for them to complete the review easily. We recommend providing external reviewers with a brief description of your review methodology, a copy of the SR plan, the main tables of evidence, and a list of all the considered studies that were eventually excluded. Having this information about the SR allows external reviewers to quickly assess the goal and easily make recommendations to the authors for final preparation of the manuscript. These recommendations may include consideration of additional articles or exclusion of articles on the tables of evidence.

## TRANSLATING CONCLUSIONS AND RECOMMENDATIONS FOR CLINICIANS AND CONSUMERS

Once the SR is published, researchers may believe their work is complete. However, if their goal is to have the findings of their review used and implemented in practice, they need to consider additional work to develop materials based on the review that are written appropriately for clinicians and consumers. Translating evidence into information products that are easily comprehended and used by target audiences is the knowledge translation process. Consumer and clinician materials can take many forms, including factsheets, workbooks, clinical practice guidelines, and patient decision-making aids and can be presented using different media including audio (eg, podcasts), video, and text (online or paper). Decisions about format and media depend on the content to be presented and the needs of the target audience; however, all materials need to be written in plain, clear language. Plain language is defined in reference to the audience. For example, what may be plain to a broad audience of family members and individuals with disabilities will be different from what is plain to an audience of physical therapists who can be expected to understand technical terms unique to their profession.

The idea that information should be presented in plain language or plain writing is increasingly accepted. For example, as of 2010, federal law<sup>26</sup> requires all federal agencies to use plain writing in every document covered by the law (ie, documents that are necessary for understanding or obtaining benefits and services or for filing taxes, and documents that explain how to comply with a federal requirement) they release or revise. In federal law, plain writing is defined as “writing that is clear, concise, well-organized, and follows other best practices appropriate to the subject or field and intended audience.”<sup>26(Section 3[3])</sup>

Plain language summaries are also required by some research journals and the Cochrane Collaboration’s SRs. Plain language is not writing that is simple or dumbed down but is writing with the intent of communicating to a specific audience and does so with straightforward grammatical structure and accurate terminology (but not jargon).<sup>27</sup> Steinberg defines it as “The writing and setting out of essential information in a way that gives a cooperative, motivated person a good chance of understanding the document at the first reading, and in the same sense that the writer meant it to be understood.”<sup>28(p3)</sup>

Plain language requires that consumers and clinicians can easily find the information they need, understand it, and act appropriately based on that understanding.<sup>29</sup> Plain language benefits both the consumer and the developer of informational resources. Because of its clarity, readers understand the material more efficiently, more people understand the message, and fewer people misunderstand, and therefore, consumers of the information can act more independently in decision making.

A plain language summary of an SR should provide adequate information to allow a reader to understand how the review was conducted and the strength of the evidence supporting any conclusions. The Cochrane Collaboration includes the following components in their plain language summaries: (1) background, (2) objectives, (3) search strategy, (4) selection criteria, (5) data collection and analysis, (6) main results, and (7) authors’ conclusions.

From a practical perspective, it can be useful to ask someone not involved in writing the SR to write a first draft of the plain language summary. SR authors are often too close to the work and may have a hard time translating their detailed knowledge into a summary. Once a draft has been written by an individual

external to the author team, SR authors should review the draft for accuracy.

## CONCLUSIONS

The evidence supporting practice in rehabilitation is less likely to be based on criterion standard research than other areas of medicine. Nevertheless, people with disabilities, their families, health care providers, and policymakers all require access to the best quality evidence available to inform decision making and practice. They also need clear guidance in how to evaluate that evidence. This unique set of circumstances increases the burden on authors of SRs. Publishing high quality SRs is critical to the field in terms of identifying gaps in research and in supporting the practice of EBP. But authors must develop strategies to focus SRs meaningfully, capture the relevant articles, evaluate the relevant evidence, and synthesize and summarize effectively. And as we describe in the Translation Conclusions and Recommendations for Consumers and Clinicians section above, a focus on knowledge translation should always be associated with an SR, which requires not only plain language summaries or translations, but also effective deployment to relevant constituents and ongoing support for implementation of any recommendations.

Our experience in conducting a number of SRs with multiple authors over the past 5 years in the areas of TBI, SCI, and burn rehabilitation has led us to respect the enormous amount of effort required to complete quality SRs. Because of the importance of SRs in rehabilitation, and the effort required to complete quality SRs, we have recommended strategies here to increase the efficiency of the process in all phases.

**Acknowledgments:** We thank Anjali Truitt, MPH, for her assistance in the preparation of this article. We also thank supplement editors Ronald Seel, PhD, Marcel Dijkers, PhD, and Mark Johnston, PhD, for their thoughtful review of this manuscript.

## References

1. Dijkers MP, Murphy SL, Krellman J. Evidence-based practice for rehabilitation professionals: concepts and controversies. *Arch Phys Med Rehabil* 2012;93(8 Suppl 2):S164-76.
2. Liberati AA, Tetzlaff J, Mulrow C, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009;62:e1-34.
3. Johnston M, Vanderheiden G, Farkas M, Rogers E, Summers J, Westbrook J. The challenge of evidence in disability and rehabilitation research and practice: a position paper. NCDDR Task Force on Standards of Evidence and Methods. Austin: SEDL; 2009.
4. Victora CG, Habicht JP, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health* 2004;94:400-5.
5. Ylvisaker M, Turkstra L, Coelho C. Behavioral and social interventions for individuals with traumatic brain injury: a summary of the research with clinical implications. *Semin Speech Lang* 2005; 26:256-67.
6. Kennedy MR, Coelho C, Turkstra L, et al. Intervention for executive functions after traumatic brain injury: a systematic review, meta-analysis and clinical recommendations. *Neuropsychol Rehabil* 2008;18:257-99.
7. Kung J, Chiappelli F, Cajulis OO, et al. From systematic reviews to clinical recommendations for evidence-based health care: validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for grading of clinical relevance. *Open Dent J* 2010;4:84-91.
8. Lewin S, Oxman AD, Lavis JN, Gruen RL, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking

- (STP) 9: assessing the applicability of the findings of a systematic review. *Health Res Policy Syst* 2009;7(Suppl 1:S9).
9. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med* 2007;4:e78.
  10. O'Mathúna DP. Critical appraisal of systematic reviews. *Int J Nurs Pract* 2010;16:414-8.
  11. Shea B, Bouter L, Peterson J, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS One* 2007;2:e1350.
  12. Shea B, Grimshaw J, Wells G, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.
  13. Teasdale G, Jennett B. Assessment and prognosis of coma after head injury. *Acta Neurochir (Wien)* 1976;34:45-55.
  14. Sampson M, McGowan J. Errors in search strategies were identified by type and frequency. *J Clin Epidemiol* 2006;59:1057-63.
  15. Johnston M, Sherer M, Whyte J. Applying evidence standards to rehabilitation research. *Am J Phys Med Rehabil* 2006;85:292-309.
  16. Schünemann H, Fretheim A, Oxman A. Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations. *Health Res Policy Syst* 2006;4:21.
  17. Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). *Aust J Physiother* 2002;48:43-9.
  18. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52:377-84.
  19. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine: how to practice and teach EBM*. Edinburgh: Churchill Livingstone; 2000.
  20. Edlund W, Grosneth G, So Y, Franklin G. *Clinical practice guidelines process manual*. St. Paul, MN: American Academy of Neurology; 2004.
  21. Dikmen S, Machamer J, Temkin N, McLean A. Neuropsychological recovery in patients with moderate to severe head injury: 2-year follow-up. *J Clin Exp Neuropsychol* 1990;12:507-19.
  22. Johnston MV, Dijkers MP. Toward improved evidence standards and methods for rehabilitation: recommendations and challenges. *Arch Phys Med Rehabil* 2012;93(8 Suppl 2):S185-99.
  23. Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. *J Clin Epidemiol* 2009;62:944-52.
  24. Hartling L. United States. Agency for Healthcare R, Quality, University of Alberta Evidence-based Practice C. Developing and testing a tool for the classification of study designs in systematic reviews of interventions and exposures. Rockville: Agency for Healthcare Research and Quality; 2010.
  25. Elamin M, Flynn D, Bassler D, et al. Choice of data extraction tools for systematic reviews depends on resources and review complexity. *J Clin Epidemiol* 2009;62:506-10.
  26. Plain Writing Act (Public Law 111-274—Oct. 13, 2010). 5 USC 301 (2010).
  27. Mazur B. Revisiting plain language. *Technical Communication* 2000;47:205-11.
  28. Steinberg ER. *Plain language: principles and practice*. Detroit: Wayne State Univ Pr; 1991.
  29. Redish JC. What is information design? *Technical Communication* 2000;47:163-6.