Spinal Cord Injury Rehabilitation Evidence: Method of the SCIRE Systematic Review

Janice J. Eng, Robert W. Teasell, William C. Miller, Dalton L. Wolfe, Andrea F. Townson, Jo-Anne Aubut, Caroline Abramson, Jane T.C. Hsieh, Sandra Connolly, Kristin Konnyu, and the SCIRE Research Team

Spinal Cord Injury Rehabilitation Evidence (SCIRE) is a synthesis of the research evidence underlying rehabilitation interventions to improve the health of people living with SCI. SCIRE covers a comprehensive set of topics. The SCIRE used a systematic and well-defined protocol to assess and synthesize the evidence. Each article was scored for its methodological quality using either the Physiotherapy Evidence Database (PEDro) scale for randomized controlled trials or the Downs and Black tool for other types of studies. Following the individual study assessment, conclusions were drawn about the accumulated studies for each topic of interest based on the levels of evidence, quality of studies, and concurring evidence. The SCIRE project was designed for health professionals to inform them of best practices. **Key words:** *evidence-based practice, interventions, knowledge translation, methods, rehabilitation, spinal cord injury, systematic review*

Janice J. Eng, PhD, BSc (PT/OT), *is a Professor, School of Rehabilitation Sciences, University of British Columbia; Scientist, Rehab Research Lab, GF Strong Rehab Centre; and Investigator, International Collaboration on Repair Discoveries (ICORD), Vancouver, Canada.*

Robert W. Teasell, MD, FRCPC, *is a Physiatrist with Parkwood Hospital, St. Joseph's Health Care; a Clinical Scientist with the Aging, Rehabilitation and Geriatric Care Program, Lawson Health Research Institute; and Professor and Chair/Chief with the Department of Physical Medicine and Rehabilitation at the Schulich School of Medicine, University of Western Ontario, London, Ontario, Canada.*

William C. Miller, PhD, OT, *is an Associate Professor, School of Rehabilitation Sciences, University of British Columbia; Scientist, Rehab Research Lab, GF Strong Rehab Centre; and Investigator, ICORD, Vancouver, Canada.*

Dalton L. Wolfe, PhD, *is an Associate Scientist, Lawson Health Research Institute, Aging Rehabilitation and Geriatric Care, and the Schulich School of Medicine, University of Western Ontario, Department of Physical Medicine and Rehabilitation, St. Joseph's Health Care London, Parkwood Hospital, London, Ontario, Canada.*

Andrea F. Townson, MD, FRCPC, *is a Clinical Assistant Professor, University of British Columbia, Division of Physical Medicine and Rehabilitation; Medical Manager SCI Rehab Program at GF Strong Rehabilitation Centre; and Investigator, ICORD, Vancouver, British Columbia.*

Jo-Anne Aubut, BA, *is a Research Assistant with the Aging, Rehabilitation and Geriatric Care Program, Lawson Health Research Institute, London, Ontario, Canada.*

Caroline Abramson, MA, *is a Research Assistant, ICORD, Vancouver, Canada.*

Jane T.C. Hsieh, MSc, *is an Associate Scientist, Lawson Health Research Institute, Aging Rehabilitation and Geriatric Care, London, Ontario, Canada.*

Sandra Connolly, BHScOT, *is an Occupational Therapist at St. Joseph's Health Care London, Parkwood Hospital, London, Ontario, Canada.*

Kristin Konnyu, BSc, *is a Graduate Student, School of Rehabilitation Sciences, University of BC, and Research Assistant, ICORD, Vancouver, Canada.*

Top Spinal Cord Inj Rehabil 2007;13(1):1–10

www.thomasland.com

doi: 10.1310/sci1301-1

The field of spinal cord injury (SCI) rehabilitation has experienced significant advances over the past decade. Neurophysiological principles founded on animal models have developed into human trials, for example, utilization of central pattern generators in training locomotion or epidural spinal stimulation to enhance recovery. Advances in microprocessors and signal processing have resulted in more portable and efficient functional electrical stimulation systems to facilitate movement. Clinical trials have started to identify the required intensity and duration of exercise required for functional gains. Randomized controlled trials (RCTs) have provided quality evidence on the effect of pharmacological interventions on function. Unquestionably, the amount of literature in SCI rehabilitation is growing. Given the expansive growth of such increasingly complex information, it is often difficult for a front-line clinician not intimately familiar with the research methods and analyses to interpret the results of a study. In addition, the interpretation is further complicated by the presence of multiple studies on an intervention, often with what appear to be conflicting messages. The Spinal Cord Injury Rehabilitation Evidence (SCIRE) project is dedicated to providing an up-to-date, accurate research synthesis about the effect of rehabilitation interventions for people with SCI.

This issue of *Topics in Spinal Cord Rehabilitation* includes six articles relevant to SCI rehabilitation clinicians (SCI inpatient rehabilitation practices, gait strategies, upper extremity reconstructive surgery, spasticity treatments, cardiovascular health, and bone health). The SCIRE used a systematic and well-defined protocol to assess and synthesize the evidence of the effects of rehabilitation interventions in SCI and is designed for health professionals to inform them of best practices. The full SCIRE compendium is available at www.icord.org/scire.

Evaluation of the Evidence

Literature search strategy

A systematic review was undertaken using multiple databases (MEDLINE/PubMed, CINAHL®, EMBASE, PsycINFO) to identify all relevant literature published from 1980 to 2006. An initial broad search was performed with five types of SCI therapies: drug therapy, radiotherapy, diet therapy, rehabilitation therapy, and therapy. Articles were included if they involved human subjects, a specific intervention, and were published in English. Exclusion criteria included the following: less than half the reported population had an SCI or no measurable outcome associated with the intervention. In the absence of supporting literature, studies with fewer than three subjects were also included. Meta-analyses, systematic reviews, and review articles were also captured. RefWorks, an on-line research management system, was used to create the database. References of all articles were reviewed for pertinent articles, which were sought through hand searching. Hand searching may provide higher rates of return than electronic searching within a particular subject area.[1]

MeSH headings were used with the key words. Key words were paired with spinal cord injury, tetraplegia, quadriplegia, or paraplegia. Specific SCI rehabilitation topics (e.g., spasticity) were identified by a multidisciplinary team of expert scientists, clinicians, consumers with SCI, and policy makers. These specific topics were searched

Table 1. Five levels of evidence

Level	Research design	Description
Level 1	Randomized controlled trial (RCT)	Randomized controlled trial, PEDro score ≥6. Includes within-subjects comparison with randomized conditions and crossover designs.
Level 2	RCT	RCT, PEDro score <6.
	Prospective controlled trial	Prospective controlled trial (not randomized)
	Cohort	Prospective longitudinal study using at least 2 similar groups with 1 group exposed to a particular condition.
Level 3	Case control	A retrospective study comparing conditions, including historical controls.
Level 4	Pre-post	A prospective trial with a baseline measure, intervention, and a posttest using a single group of subjects.
	Posttest	A prospective posttest with 2 or more groups: intervention, then posttest (no pretest or baseline measurement) using a single group of subjects.
	Case series	A retrospective study usually collecting variables from a chart review.
Level 5	Observational	Study using cross-sectional analysis to interpret relations.
	Clinical consensus	Expert opinion without explicit critical appraisal, or based on physiology, biomechanics, or "first principles."
	Case report	Pre-post or case series involving 1 subject.

Note: PEDro = Physiotherapy Evidence Database scale.

with additional key words generated from expert scientists and clinicians in SCI rehabilitation familiar with the topic, and more titles and abstracts were reviewed. The specific key words utilized for the topics reviewed in this issue are listed in the SCIRE compendium.[2] The search involved the review of over 17,000 titles and 8,400 abstracts and a final extraction and synthesis of almost 700 articles.

Quality assessment tools and data extraction

The research design of the study was initially determined (e.g., RCT; see **Table 1** for full list of research designs). Next, each article was scored for its methodological quality using either the Physiotherapy Evidence Database (PEDro) scale[3] or the Downs and Black tool.[4] Two independent raters reviewed each article and determined the qual-

ity score. Intraclass correlations (ICC [model 1,K]) for the two independent scores were 0.91 (95% CI 0.70–0.98) for the PEDro score (*F* statistic not significant) and 0.90 (95% CI 0.88–0.92) for the Downs and Black tool (*F* statistic not significant). Scoring discrepancies were resolved through discussion. Tables were generated from the extracted data, which included sample subject characteristics, nature of the intervention, outcome measures, and key results.

The PEDro score[3] was used to assess the methodological quality of individual RCTs. PEDro was developed for the purpose of accessing bibliographic details and abstracts of RCTs, quasi-randomized studies, and systematic reviews in physiotherapy. PEDro has been used to assess both pharmacological and nonpharmacological studies with good agreement between raters at an individual item level and in total PEDro scores.[5] Maher et al.[6] found the reliability of PEDro item ratings varied from "fair" to "substantial," whereas the reliability of the total PEDro score was "fair" to "good." The PEDro is an 11-item scale, in which the first item relates to external validity and the other 10 items assess the internal validity of a clinical trial. One point is given for each satisfied criterion (except for the first item, which is given a "yes" or "no"), yielding a maximum score of 10. The higher the score, the better the quality of the study as assessed by the following cut-points defined by Foley et al.[7] : 9–10, *excellent*; 6–8, *good*; 4–5, *fair*; <4, *poor*. A point for a particular criterion was awarded only if the article explicitly reported that the criterion was met. The scoring system is detailed in **Appendix 1**.

All other studies with an intervention were assessed with the Downs and Black tool[4] for methodological quality. Saunders et al.[8] compared 18 instruments that assessed the quality of nonrandomized trials. From their analyses, they suggested that the Downs and Black tool was the best because it was developed using psychometric methods and had been tested for reliability and validity. This tool consists of 27 questions in the following subsections: reporting, external validity, internal validity (bias), and internal validity (confounding). The original tool ranged from 0 to 32. However, the wording for the last question was ambiguous and difficult to score. Thus, we modified the last question from a scale of 0 to 5 to a scale of 0 to 1, where 1 was scored if a power calculation or sample size calculation was present and 0 was scored if there was no power calculation, sample size calculation, or explanation whether the number of subjects was appropriate. Thus, our modified version ranged from 0 to 28, with a higher score indicating higher methodological quality. This modified Downs and Black tool is attached in **Appendix 2**.

Determining levels of evidence and formulating conclusions

After the individual study assessment, conclusions were drawn about the accumulated studies for each topic of interest (e.g., effect of treadmill training on gait outcomes) based on the levels of evidence developed by Sackett et al.[9] The subcategories with each level (e.g., level 1a, 1b, 1c) proposed by Sackett et al.[9] were collapsed into a single level. This was performed to reduce the 10 subcategories from Sackett et al.[9] to a less complex system from level 1 to level 5 that included the key research designs encountered in the SCI rehabilitation literature (**Table 1**). We provided additional descriptions specific to the types of research designs found in SCI rehabilitation to facilitate the decision-making process.

Sackett et al.[9] distinguishes high- and low-quality RCTs into level 1b and level 2b, respectively. However, the specific criteria for this distinction are not provided. To provide a more reliable decision-making process, we required that a level 1 RCT have a PEDro score of greater than or equal to 6 (good to excellent quality), whereas a level 2 RCT have a PEDro score of 5 or less. Note that we report sample sizes so the reader is aware of the numbers of subjects on which the conclusions are based. There was no sample size requirement for level 1 evidence, but these studies were rated highly according to the standardized PEDro scale. The appropriateness of the control group was assessed per study. In some studies, an able-bodied group may not have been an adequate control for the particular intervention used but simply provided "normative" values for comparison. In those studies, the study was considered "not controlled" and the level of evidence reduced (e.g., level 4 pre-post).

RCTs received priority when formulating conclusions. If studies addressing the same treatment differed in quality, more weight was applied to the studies with higher quality scores when deriving the final conclusions. Although conclusions were based on the study designs, levels of evidence, and concurring evidence, there were some difficult instances, such as the results of a single high-quality study conflicting with those of several lower quality studies. When conflicting data were present, an explanation was provided as to how the conclusions were derived to make the process as transparent as possible.

Concluding Remarks

The SCIRE provides a synthesis of the research evidence underlying rehabilitation interventions that is transparent and reproducible. However, it is not without its limitations. Interventions published in the literature may not exactly match what is currently practiced in the clinical setting. As emphasized by Sackett et al.,[10] the evidence from systematic research should be integrated with clinical expertise and patients' choice to form best practice. For example, the extensive literature on functional electrical stimulation in spinal cord rehabilitation is not matched by its practice in the clinical or community setting; this could be due to a variety of reasons, such as lack of uptake of practice by clinicians, clinician and patient preferences and priorities, lack of an inexpensive commercial product, physical difficulties in using the system, or excessive cost (monetary expense, time for treatment, cosmetic acceptance) for the resulting benefits in function.[11–13]

SCIRE allows for objective identification of areas in which there is substantial quality evidence for practice, as well as the gaps in which additional research is urgently required. Although the conclusions reached by SCIRE may change as additional studies are undertaken, new evidence can be incorporated with the same rigor with minimal bias.

Acknowledgments

We acknowledge the financial support for the SCIRE project from the Rick Hansen Man-in-Motion Foundation and Ontario Neurotrauma Fund. In addition, salary support was provided from the Canadian Institutes of Health Research (to Drs. Eng and Miller) and the Michael Smith Foundation for Health Research (to Dr. Eng).

REFERENCES

1. Hopewell S, Clarke M, Lefebvre C, Scherer R. Handsearching versus electronic searching to identify reports of randomized trials. *The Cochrane Database of Methodology Reviews.* 2002; Issue 4. Art. No. MR000001.
2. Methods of the systematic reviews. In: Eng JJ, Teasell RW, Miller WC, et al., eds. *SCIRE: Spinal Cord Injury Rehabilitation Evidence.* 2006:2.1–2.11. Available at: http://www.icord.org/scire.
3. Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). *Aust J Physiother.* 2002;48:43–49.
4. 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;52:377–384.
5. Foley NC, Bhogal SK, Teasell RW, Bureau Y, Speechley MR. Estimates of quality and reliability with the physiotherapy evidence-based database scale to assess the methodology of randomized controlled trials of pharmacological and nonpharmacological interventions. *Phys Ther.* 2006;86:817–824.
6. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. *Phys Ther.* 2003;83:713–721.
7. Foley NC, Teasell RW, Bhogal SK, Speechley MR. Stroke Rehabilitation Evidence-Based Review: methodology. *Top Stroke Rehabil.* 2003;10(1):1–7.
8. Saunders LD, Soomro GM, Buckingham J, Jamtvedt G, Raina P. Assessing the methodological quality of nonrandomized intervention studies. *West J Nurs Res.* 2003;25:223–237.
9. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-Based Medicine: How to Practice and Teach EBM.* Toronto, Ontario: Churchill Livingstone; 2000.
10. Sackett DL, Rosenberg WM, et al. Evidence based medicine: what it is and what it isn't. *BMJ.* 1996;312:71–72.
11. Brissot R, Gallien P, Le Bot MP, Beaubras A, Laisne D, Beillot J, Dassonville J. Clinical experience with functional electrical stimulation-assisted gait with Parastep in spinal cord-injured patients. *Spine.* 2000;25:501–508.
12. Ditunno PL, Patrick M, Stineman M, Morganti B, Townson AF, Ditunno JF. Cross-cultural differences in preference for recovery of mobility among spinal cord injury rehabilitation professionals. *Spinal Cord.* 2006;44:567–575.
13. Estores IM. The consumer's perspective and the professional literature: what do persons with spinal cord injury want? *J Rehabil Res Dev.* 2003;40(4 suppl 1):93–98.

APPENDIX 1

The PEDro Score

1. "Subjects were randomly allocated to groups." In a crossover study, subjects were randomly allocated an order in which treatments were received. A point for random allocation was awarded if random allocation of patients was stated in its methods. The precise method of randomization need not be specified. Procedures such as coin-tossing and dice-rolling were considered random. Quasi-randomization allocation procedures such as allocation by bed availability did not satisfy this criterion.
2. "Allocation was concealed." A point was awarded for concealed allocation if this was explicitly stated in the methods section or if there was reference that allocation was by sealed opaque envelopes or that allocation involved contacting the holder of the allocation schedule who was "off-site."
3. "The groups were similar at baseline regarding the most important prognostic indicators." A trial was awarded a point for baseline comparability if at least one key outcome measure at baseline was reported for the study and control groups. This criterion was satisfied even if only baseline data of study completed-only subjects were presented.
4. "There was blinding of all subjects." The person in question (subject, therapist, or assessor) was considered blinded if he/she did not know to which group the subject had been allocated. In addition, subjects and therapists were only considered to be "blind" if it could be expected that they would have been unable to distinguish between the treatments applied to different groups. In drug therapy trials, the administrator of the drug was considered the therapist and was considered blinded if he/she did not prepare the drug and was unaware of the drug being administered.
5. "There was blinding of all therapists who administered the therapy." Criteria 4.
6. "There was blinding of all assessors who measured at least one key outcome." Criteria 4.
7. "Adequacy of follow-up." For the purposes of this review, follow-up was considered adequate if all of the subjects who had been originally randomized could be accounted for at the end of the study. The interpretation of this criterion differs from that described by PEDro, where adequacy is defined as the measurement of the main outcome in more than 85% of subjects.
8. "Intention to treat." All subjects for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome were analyzed by "intention to treat." For purposes of the present evidence-based review, a trial was awarded a point for intention-to-treat if the trial explicitly stated that an intention-to-treat analysis was performed.
9. "The results of between-group statistical comparisons are reported for at least one key outcome." Scoring of this criterion was design dependent. As such, between-group

comparison may have involved comparison of two or more treatments or comparison of treatment with a control condition. The analysis was considered a between-group analysis if either a simple comparison of outcomes measured after the treatment was administered was made or a comparison of the change in one group with the change in another was made. The comparison may be in the form of hypothesis testing (e.g., *p* value) or in the form of an estimate (e.g., the mean, median difference, difference in proportion, number needed to treat, relative risk, or hazard ratio) and its confidence interval. A trial was awarded a point for this criterion if between-group comparison on at least one outcome measure was made and its analysis of comparison was provided.

10. "The study provides both point measures and measures of variability for at least one key outcome." A point measure was referred to as the measure of the size of the treatment effect. The treatment effect was described as being either a difference in group outcomes or as the outcome in (each of) all groups. Measures of variability included standard deviations, standard errors, confidence intervals, interquartile ranges (or other quartile ranges), and ranges. Point measures and/or measures of variability that were provided graphically (for example, *SD*s may be given as error bars in a figure) were awarded a point as long as it was clear what was being graphed (e.g., whether error bars represent *SD*s or *SE*s). For those outcomes that were categorical, this criterion was considered to have been met if the number of subjects in each category was given for each group.

Adapted, with permission, from Moseley AM, Herbert RD, Sherrington C, Maher CG. Evidence for physiotherapy practice: a survey of the Physiotherapy Evidence Database (PEDro). *Aust J Physiother.* 2002;48:43–49. Copyright © 2002 by the Australian Physiotherapy Association.

APPENDIX 2

Modified Downs and Black tool

Reporting

1. Is the hypothesis/aim/objective of the study clearly described?
2. Are the main outcomes to be measured clearly described in the Introduction or Methods section?
3. Are the characteristics of the patients included in the study clearly described?
4. Are the interventions of interest clearly described?
5. Are the distributions of principal confounders in each group of subjects to be compared clearly described?
6. Are the main findings of the study clearly described?
7. Does the study provide estimates of the random variability in the data for the main outcomes?
8. Have all important adverse events that may be a consequence of the intervention been reported?
9. Have the characteristics of patients lost to follow-up been described?
10. Have actual probability values been reported (e.g., .035 rather than <.05) for the main outcomes except where the probability value is less than .001?

External validity

11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?
12. Were those subjects who were prepared to participate representative of the entire population from which they were recruited?
13. Were the staff, places, and facilities where the patients were treated representative of the treatment the majority of patients receive?

Internal validity – bias

14. Was an attempt made to blind study subjects to the intervention they received?
15. Was an attempt made to blind those measuring the main outcomes of the intervention?
16. If any of the results of the study were based on "data dredging," was this made clear?
17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients or in case-control studies is the time period between the intervention and outcome the same for cases and controls?
18. Were the statistical tests used to assess the main outcomes appropriate?
19. Was compliance with the intervention(s) reliable?
20. Were the main outcome measures used accurate (valid and reliable)?

Internal validity – confounding (selection bias)

21. Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?
22. Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?
23. Were study subjects randomized to intervention groups?
24. Was the randomized intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?
25. Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?
26. Were losses of patients to follow-up taken into account?
27. Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance was less than 5%?

Modified, with permission, from 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;52:377–384.