EDITORIAL
Improving the quality of systematic reviews in *Spinal Cord*

Systematic reviews about treatment effectiveness are increasingly abundant in the area of spinal cord injuries (SCI). They are important because they summarise the evidence upon which clinical practice guidelines are based. Given the abundance of systematic reviews in the area of SCI, it is perhaps now appropriate that *Spinal Cord* works with authors to further improve the quality of our systematic reviews.

One obvious way to improve the quality of our systematic reviews is to ensure they adhere to the PRISMA Statement for the reporting of systematic reviews.[1] The PRISMA Statement contains a 27-item checklist covering all important methodological aspects of systematic reviews. The PRISMA Statement emphasises the importance of ensuring systematic reviewers are driven by clearly articulated clinical questions. This involves defining the types of interventions and outcomes of interest within formal protocols before commencing systematic reviews. The protocols need to state all important decisions and be either published or registered.[2] It is rarely appropriate to merely search for all trials related to a broad topic and upon completion set about categorising the results of the identified trials. This later approach is vulnerable to bias from selective reporting and opportunistic extraction of data.

Another important way authors can improve the quality of their systematic reviews is by ensuring they report estimates of the size and precision of treatment effects (for example, as between-group differences or odds ratios with their confidence intervals). This should be done regardless of whether meta-analyses are performed. Of course, the authors of the original research do not always report their results in this format. However, it is nearly always possible to mathematically manipulate the reported data to get the results in the correct format. (The Cochrane Collaboration provides extensive and freely available online resources to help the authors of systematic reviews do this.) Authors then need to look closely at the size and precision of treatment estimates. Very small treatment effects need to be questioned for clinical relevance, especially if only surrogate outcomes are used. Very large treatment effects, even those reported in low-quality trials, may be important, provided the results are consistent across many different sources.[3] The confidence intervals associated with between-group differences also need to be examined because they provide an indication of how confident we can be about estimates of treatment effects. Quality of evidence should be downgraded if the width of the confidence interval about an estimate of treatment effect is wide and if the confidence interval crosses a decision threshold.[4] This approach to data analysis and interpretation is far superior to merely relying on *P* values of identified trials to make judgements about treatment effectiveness.[5] *P* values give no indication of the size or precision of estimates of treatment effects and may be uninformative and misleading. It is particularly problematic solely relying on *P* values associated with within-group differences (that is, pre to post changes) with no direct comparison of these differences between groups. This later problem cannot be adequately overcome by merely reporting the size of pre to post changes for each group regardless of how these changes are expressed. Reporting percentage changes for each group and then implying a comparison between groups is notoriously misleading. Authors must resist this type of superficial reporting of results in systematic reviews and ensure all outcomes are expressed as contrasts between groups with their confidence intervals, and interpret the results accordingly.

Rating evidence in Systematic Reviews is not merely a matter of rating the quality of the included studies. There are many additional subtleties, which need to be taken into account. One method for rating evidence is the GRADE method (an acronym for, Grading of Recommendations, Assessment, Development and Evaluation). This is advocated by organisations such as the Cochrane Collaboration, BMJ, the American College of Physicians and the World Health Organisation.[6–8] The GRADE method requires the authors to identify outcomes that are of key importance to patients and discourages the authors from relying on surrogate outcomes. The evidence supporting the effectiveness of an intervention on a particular outcome is then rated on a 4-point scale ranging from 'very low' to 'high'. A number of different factors are taken into account when providing ratings. These include five factors that can lower our confidence in estimates of effect (that is, risk of bias, inconsistency of results across studies, indirectness of the evidence, imprecision of estimates and publication bias) and three factors that can increase our confidence (that is, large effects, a dose–response relationship and effects that are opposite to what would be expected from the influences of confounding and bias).[1,6–8,9] Freely available software guides the authors through each of these judgements.(http://ims.cochrane.org/revman/gradepro).

As *Spinal Cord* moves forward and continues to advocate evidence-based care, we will encourage authors to adhere to the PRISMA Statement and to report the size and precision of treatment estimates. We will also encourage authors to consider using the GRADE method to rate the evidence. After all, a high-quality systematic review that provides a robust estimate of treatment effectiveness and a thorough summary of the evidence is far more valuable than any number of low-quality reviews that may only serve to derail evidence-based care.

CONFLICT OF INTEREST

The author declares no conflict of interest.

LA Harvey

Rehabilitation Studies Unit, Northern Clinical School, Sydney School of Medicine, University of Sydney, Sydney, NSW, Australia
E-mail: l.harvey@usyd.edu.au

1 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; **6**: e1000097.
2 Booth A, Clarke M, Ghersi D, Moher D, Petticrew M, Stewart L. An international registry of systematic-review protocols. *Lancet North Am Ed* 2011; **377**: 108–109.

3 Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P *et al.* GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 2011; **64**: 1311–1316.

4 Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D *et al.* GRADE guidelines 6. Rating the quality of evidence - Imprecision. *J Clin Epidemiol* 2011; **64**: 1283–1293.

5 Goodman S. Toward evidence-based medical statistics. 1: the *P* value fallacy. *Ann Intern Med* 1999; **130**: 995–1004.

6 Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A *et al.* Going from evidence to recommendations. *BMJ* 2008; **336**: 1049–1051.

7 Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ. What is "quality of evidence" and why is it important to clinicians? *BMJ* 2008; **336**: 995–998.

8 Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; **336**: 924–926.

9 Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J *et al.* GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011; **64**: 383–394.