Systematic reviews and clinical practice guidelines are needed to inform and guide clinical practice in physiotherapy. Clinical practice guidelines should be based on systematic reviews, and both systematic reviews and clinical practice guidelines should rate the quality of evidence. However, only clinical practice guidelines should make direct recommendations about clinical practice because recommendations depend on information and judgements that go beyond systematic reviews (Guyatt et al 2008a).

Many systematic reviews and clinical practice guidelines rate the strength of evidence primarily on the basis of study design, risk of bias, and reported p values. For example, evidence from randomised controlled trials that report statistically significant findings is rated highly. Similarly, randomised controlled trials that conceal allocation, blind assessors, and minimise drop outs are rated higher than trials that do not. This approach ignores many important aspects of evidence that need to be taken into account when rating its quality. For example, it ignores how confident we are in an estimate of the effect of a therapy and the relative importance of different types of outcomes to people who seek physiotherapy interventions. In addition, a sole focus on p values ignores imprecision which should be used to downgrade the quality of evidence and ignores other factors that can either decrease or increase our confidence in estimates of effect. Given the abundance of systematic reviews and the growing number of clinical practice guidelines, it is perhaps now appropriate that the international physiotherapy community focuses on improving the way we rate evidence in our reviews and guidelines.

One way to improve the way we rate evidence in our systematic reviews and clinical practice guidelines is to fall in line with organisations such as BMJ Group, the Cochrane Collaboration, the American College of Physicians and the World Health Organisation, and use the GRADE system (Guyatt et al 2008a, Guyatt et al 2008b, Guyatt et al 2008c). The GRADE system (an acronym for Grading of Recommendations Assessment, Development and Evaluation) was first published in 2004. It requires authors to initially identify outcomes that are of key importance to patients and discourages authors from relying on surrogate outcomes. The evidence supporting the effectiveness of an intervention on an outcome is then rated on a 4-point scale ranging from ‘very low’ to ‘high’.

Notably, evidence about the effectiveness of interventions on each outcome is not just rated according to study design or p values, although these are considered. Instead, evidence is also rated according to a number of factors. These include five factors that can lower our confidence in estimates of effect (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 (large effects, a dose response relationship, and effects that are opposite to what would be expected from the influences of confounding and bias). Freely available software (GRADEpro) can guide authors through each of these judgements. Some judgements are easier and less ambiguous to make than others. However, all important factors that influence our confidence in estimates of the effect of an intervention are taken into account when rating the strength of the evidence.

Two key factors taken into account by the GRADE system are the size and precision of estimates. The precision of estimates is reflected in the width of confidence intervals and tells us how confident we can be in an estimate. Quality of evidence should be downgraded if the width of the confidence interval for an estimate of treatment effect is large and if the confidence interval crosses a decision threshold (Guyatt et al 2011a). Similarly, the size of treatment effects is an important consideration. Observational studies that indicate very large treatment effects can provide moderate or even high quality evidence for an intervention. Although observational studies often overestimate treatment effects due to confounding, this alone cannot explain very large treatment effects (Guyatt et al 2011b). Consideration of the size and precision of estimates requires moving beyond p values, which may be misleading and are often misinterpreted (Goodman 1999). There are of course many other subtleties involved in using the GRADE system to rate the quality of evidence and readers are referred to the many excellent, freely available resources (eg, see Guyatt et al 2008a, Guyatt et al 2008b, Guyatt et al 2008c, Guyatt et al 2011c).

As the international physiotherapy community moves forward and continues to advocate for evidence-based care, we should be encouraging authors of systematic reviews and clinical practice guidelines to use the GRADE system to rate the quality of evidence in their systematic reviews and clinical practice guidelines, and the strength of recommendations in guidelines. Importantly, we should be encouraging better reporting of original comparative research to help authors of reviews and clinical practice guidelines adopt the GRADE system. For example, authors of original research need to clearly articulate, where appropriate, mean between-group differences and 95% CIs for continuous data, and risk ratios and 95% CIs for dichotomous data. Journal of Physiotherapy will continue to advocate for the adoption of GRADE and better reporting of comparative research in its efforts to help advance evidence-based physiotherapy.

References

Websites
GRADEpro: ims.cochrane.org/gradepro
GRADEpro help: www.gradeworkinggroup.org/FAQ/index.htm