The quality of reporting might not reflect the quality of the study: implications for undertaking and appraising a systematic review

Chris Littlewood¹, Jon Ashton², Ken Chance-Larsen³, Stephen May³, Ben Sturrock³

¹University of Sheffield, UK, ²Doncaster & Bassetlaw Hospitals NHS Trust, UK, ³Sheffield Hallam University, UK

The systematic review has become an increasingly popular method of synthesizing findings on a topic in order to inform clinical practice, commissioning of care, and future research. A central component of the systematic review is an assessment of study quality or risk of bias, i.e. an assessment of how near to the 'truth' the findings of the study are. While undertaking a recent systematic review, it became apparent that the outcomes of the quality appraisal process were somewhat different across systematic reviews where the same randomized controlled trials had been included. The quality of the report writing of the randomized controlled trials included was identified as one possible reason for this discrepancy. This had implications upon the conclusions drawn by the review. It is suggested that reasonable attempts to contact study authors should be made in order to inform the quality appraisal process while undertaking systematic reviews of randomized controlled trials and that the presence or absence of this process should be considered by research consumers when appraising the quality of a systematic review. This process enables a full assessment of study quality rather than simply an assessment of the quality of report writing.

Keywords: Systematic review, Research methodology, Consort statement

Introduction

Over recent years, there has been a dramatic rise in the volume of published literature which makes it almost impossible for relevant stakeholders, including patients, clinicians, researchers, and commissioners, to keep abreast of developments in physiotherapy and related fields.¹ The systematic review is one approach that has been designed to address this difficulty with the aim of synthesizing the findings of primary studies on a single topic in order to assess the overall clinical impact and relevance of that body of literature.² This popular research method utilizes systematic and transparent means to identify, select, appraise, and synthesize research and frequently underpins national and international practice guidelines.^{3,4}

A key component of the systematic review process is an assessment of study quality or risk of bias, i.e. an assessment of how near to the 'truth' the findings of the study are.⁵ Most systematic reviews now include this before drawing conclusions.³ This is important because previous studies have suggested

that inadequate conduct of randomized controlled trials (RCTs) including the methods of randomization, concealment of allocation, and blinding might lead to an exaggeration of treatment effect.⁶ Hence, an introduction of bias through inadequate research methods might impact upon conclusions drawn through the systematic review process.

Most systematic reviews rely upon published reports to extract data and subsequently appraise study quality. This is despite previous literature identifying discrepancies between published reports and the actual conduct of studies.⁶⁻⁸ The CONSORT statement, initially published in 1996, was developed to address inadequate reporting of RCTs and offers clear guidance for authors.⁹ Most journals now require that reports are prepared with this statement in mind.

However, while recently undertaking a systematic review, it became evident that the outcome of the study appraisal process was different to other systematic reviews where the same studies had been included. The following discussion is a reflection upon this process which concludes with recommendations for undertaking and appraising systematic reviews in physiotherapy.

Correspondence to: Chris Littlewood, University of Sheffield, Sheffield, UK. Email: c.littlewood@sheffield.ac.uk

Methods

A systematic review of RCTs was undertaken to evaluate the effectiveness of exercise for rotator cuff tendinopathy. A total of five published reports describing four studies were included.^{10–14} This review was undertaken in accordance with published guidelines by a team of reviewers with more than one member involved at each stage to minimize bias.

The risk of bias of the included studies was undertaken using the Cochrane Back Review Group (CBRG) risk of bias tool.¹⁵ It has been recognized that this tool is also useful for the assessment of trials in conditions other than back pain, hence its use in our review.¹⁶ Each item was rated as yes (=1), no (=0), and unclear (=0). A study with a low risk of bias was defined as one fulfilling six or more of the criteria items and with no fatal flaw which is defined as:

1. drop-out >50%;
2. statistically and clinically significant differences between groups at baseline indicating unsuccessful randomization.

In terms of estimating the risk of bias and hence the potential for exaggerating effect size, this approach has been validated.¹⁶ All of the studies included in our review were regarded as presenting a low risk of bias in accordance with the CBRG process and guidance.

As part of the data extraction and subsequent quality appraisal process, the authors of the included RCTs were contacted for clarification of methodological issues where information offered in the article was unclear. Contact was made via e-mail either through the last e-mail address presented in a published paper or via a name search using the Internet search engine Google.

Also, during the developmental stages of our review, several similar systematic reviews were identified.^{17–23} Upon completion of the review, we cross-referenced our findings with that of the others and at this stage discrepancies became clear. It is ineffectual to compare the outcome of quality appraisal undertaken within systematic reviews by looking at quality scores alone because different systematic reviews tend to use different quality tools upon which to base their assessment. This lack of a consensus reflects the lack of a current gold standard, but many quality appraisal tools designed to evaluate intervention studies do include similar individual criteria and it is here where the focus of this paper will be.

Results

While undertaking a systematic review, the authors of the included RCTs were contacted to seek clarification upon methodological issues that were not clear from the published study report. This resulted in an improvement of the quality rating of all studies when

clarification was received from these authors. This change in score on the CBRG risk of bias tool ranged from 2^{10–13} to 4¹⁴ points.

Subsequently, the outcome of our quality appraisal process was compared with the other seven systematic reviews which had been identified. Our systematic review utilized the CBRG risk of bias tool, whereas the others utilized the PEDro scale,^{18,19,22} MacDermid's evaluation guidelines for determining the quality of an intervention study,²¹ and tools apparently developed by the authors,^{17,20} including one that was based upon accepted methods for design and conduct of intervention research at the time of publication.²³

Table 1 illustrates how the systematic reviews identified compared the included studies across criteria common to the review criteria undertaken in our review. One of the RCTs¹² included in our review was not included in any other review and was therefore omitted from the comparison table. Two of the identified systematic reviews^{18,21} only report overall quality scores and not the breakdown of how the individual criteria were met. Both corresponding authors of these systematic reviews were contacted for further information. Data were obtained from Kelly *et al.*,¹⁸ but Michener *et al.*²¹ had disposed of the detail of the quality appraisal process and so this was not available for further analysis.

Table 1 identifies multiple discrepancies between the systematic reviews relating both to disagreement of the presence or absence of a design feature as well as disagreement due to insufficient information presented in the published reports.

Discussion

This study has identified that the quality scores of RCTs included in a systematic review increased when methodological clarification was sought from the authors of those RCTs. Furthermore, differences in the reporting of study quality between systematic reviews that have included similar studies have been highlighted. Although different quality appraisal tools have been utilized, the common features of each of these tools have repeatedly been reported differently by the authors of this group of reviews. These discrepancies relate to both disagreement of the presence or absence of a design feature and disagreement due to insufficient information presented in the published reports. Two of the identified systematic reviews^{18,21} only report overall quality scores and not the breakdown of how the individual criteria were met. Considering these identified discrepancies, this lack of transparency limits the capacity of the reader to fully appraise the systematic review process and hence make a judgement about the validity of its conclusions.

Table 1 Comparative quality appraisals (unmarked cells indicate that the corresponding criteria were not assessed during the related systematic review)

Study	Systematic review	Quality criteria											
		Was the method of randomization adequate?	Was the treatment allocation concealed?	Was the patient blinded to the intervention?	Was the care giver blinded to the intervention?	Was the outcome assessor blinded to the intervention?	Was the drop-out rate described and acceptable?	Were all randomized participants analysed in the group to which they were allocated?	Free of selective reporting?	Similarity of baseline characteristics?	Co-interventions avoided or similar?	Compliance acceptable?	Timing of outcome assessments similar?
Brox <i>et al.</i> (1993) ¹⁰	Littlewood <i>et al.</i>	✓	✓	-	-	✓	✓	✓	✓	✓	-	✓	✓
	Dorrestijn <i>et al.</i> (2009) ¹⁷	✓	?			✓	✓	✓	-/✓	✓			
	Kelly <i>et al.</i> (2010) ¹⁸	✓	-	-	-	✓	✓	✓	✓				
	Kromer <i>et al.</i> (2009) ¹⁹	✓	-	-	-	✓	✓	✓	✓				
	Kuhn (2009) ²⁰	✓				✓	✓						
Brox <i>et al.</i> (1999) ¹¹	van der Heijden <i>et al.</i> (1997) ²³	?	?	-	-	-							
	Littlewood <i>et al.</i>	✓	✓	-	-	-	✓	✓	✓	✓	-	✓	✓
	Kelly <i>et al.</i> (2010) ¹⁸	✓	-	-	-	✓	-	✓	✓				
	Kromer <i>et al.</i> (2009) ¹⁹	✓	-	-	-	✓	✓	✓	✓				
	Kuhn (2009) ²⁰	✓				✓	✓						
Ludewig and Borsiad (2004) ¹³	Michener <i>et al.</i> (2004) ²¹												
	Littlewood <i>et al.</i>	✓	✓	-	-	-	✓	✓	✓	✓	✓	?	✓
	Kelly <i>et al.</i> (2010) ¹⁸												
	Kromer <i>et al.</i> (2009) ¹⁹	✓	✓	-	-	-	✓	✓		-			
	Kuhn (2009) ²⁰	✓				?	✓						
	Michener <i>et al.</i> (2004) ²¹												
	Trampas and Kitsios (2006) ²²	✓	✓	-	-	✓	✓	✓	✓	✓			

Table 1 Continued.

		Quality criteria												
Study	Systematic review	Were all randomized participants analysed in the group to which they were allocated?											Timing of outcome assessments similar?	
		Was the method of randomization adequate?	Was the treatment allocation concealed?	Was the patient blinded to the intervention?	Was the care giver blinded to the intervention?	Was the outcome assessor blinded to the intervention?	Was the drop-out rate described and acceptable?	Free of selective reporting?	Similarity of baseline characteristics?	Co-interventions avoided or similar?	Compliance acceptable?			
Walther <i>et al.</i> (2004) ¹⁴	Littlewood <i>et al.</i>	✓	✓	✓	–	✓	✓	✓	✓	✓	✓	✓	✓	
	Kelly <i>et al.</i> (2010) ¹⁸	✓	–	–	–	–	–	–	–	–	–	–	–	
	Kromer <i>et al.</i> (2009) ¹⁹	✓	–	–	–	–	✓	✓	✓	✓	✓	✓	✓	
	Kuhn (2009) ²⁰	?				?	✓	✓						
	Michener <i>et al.</i> (2004) ²¹													
	Trampusas and Kitsios (2006) ²²	✓	?	–	–	✓	✓	?	✓	✓	✓	✓	✓	

It seems that the identification of these discrepancies is important with respect to the conduct of future systematic reviews and the appraisal of current reviews. The findings of our review regarding the role of exercise for rotator cuff tendinopathy were generally more optimistic than those of other similar systematic reviews.^{17-19,21,23} We have suggested that one reason for this could be the more specific inclusion criteria relating to study population, i.e. rotator cuff tendinopathy, rather than the more generic term 'sub-acromial impingement', the intervention, i.e. exercise incorporating a loading strategy, and study type, i.e. RCTs only to minimize the impact of bias associated with other study types.

However, another reason for the discrepancy could relate to how we regarded the risk of bias or quality of the included studies. Without failure, the risk of bias or quality rating of all studies improved when clarification was received from the authors of the studies. With regard to one study,¹⁴ clarification was sought in relation to the method of randomization, allocation concealment, patient blinding, and timing of outcome assessments which amounted to a change in quality score of four points. This compares to another review where the study was rated as only 3/10 on the PEDro scale.¹⁸ Thus, by incorporating contact with study authors into the systematic review methodology, a full assessment of the risk of bias or study quality can be undertaken rather than just an assessment of the quality of the report writing.

A recent hand search of one physiotherapy specific journal identified that 14 systematic reviews had been published between 2008 and 2011. Of these 14 systematic reviews, only two or 14% clearly indicated that attempts had been made to contact the authors of the included studies but without further comment as to the outcomes of this process or addition to the rigour of the systematic review. Of the reviews that did not approach study authors, some reported insufficient information in the published reports of included studies as a barrier to making an informed judgement about methodological quality but did not appear to attempt to address this. Despite this limitation in systematic review methodology, it was not uncommon for the systematic reviewers to conclude that further well-designed studies are needed. In some situations, such a conclusion might be challenged if a full appraisal of the included studies was undertaken rather than just an appraisal of the quality of report writing. The findings of this study have implications for previous reviews that have not carried out this process which might subsequently be regarded as being at risk of misrepresenting the strength of the available evidence. This is particularly important with respect to physiotherapy where it is not uncommon for

systematic reviews to include small number of studies and a misinterpreted study might significantly bias the conclusions of that review.

Conclusion

Inadequate reporting of RCTs included in systematic reviews might not reflect the actual quality of the methods employed. Contacting study authors clearly needs to be integral to assessing the quality of RCTs. Among other things, the absence of this process should make the reader cautious when interpreting the findings of published systematic reviews.

References

- 1 Carroll LJ, Cassidy JD, Peloso PM, Giles-Smith L, Cheng CS, Greenhalgh SW, *et al.* Methods for the best evidence synthesis on neck pain and its associated disorders: the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders. *Spine*. 2008;5:S33–8.
- 2 Assendelft W, Koes B, Knipschild P, Bouter L. The relationship between methodological quality and conclusions in reviews of spinal manipulation. *JAMA*. 1995;274:1942–8.
- 3 Higgins J, Green S. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley Blackwell; 2008.
- 4 Hettinga D, Hurley D, Jackson A, May S, Mercer C, Roberts L. Assessing the effect of sample size, methodological quality and statistical rigour on outcomes of randomised controlled trials on mobilisation, manipulation and massage for low back pain of at least 6 weeks duration. *Physiotherapy*. 2008;94:97–104.
- 5 Centre for Reviews & Dissemination (CRD). *Systematic reviews: CRD's guidance for undertaking reviews in health care*. Hestington: CRD, University of York; 2008.
- 6 Hill CL, LaValley MP, Felson DT. Discrepancy between published reports and actual conduct of randomized clinical trials. *J Clin Epidemiol*. 2002;55:783–6.
- 7 Devereaux PJ, Choi P, El-Dika S, Bhandari M, Montori VM, Schunemann HJ, *et al.* An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. *J Clin Epidemiol*. 2004;57:1232–6.
- 8 Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, *et al.* Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ*. 2004;328:22–4.
- 9 Schulz K, Altman D, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:698–702.
- 10 Brox J, Staff P, Ljunggren A, Brevik J. Arthroscopic surgery compared with supervised exercises in patients with rotator cuff disease (stage II impingement syndrome). *BMJ*. 1993;307:899–903.
- 11 Brox J, Gjengedal E, Uppheim G, Bohmer A, Brevik J, Ljunggren A, *et al.* Arthroscopic surgery versus supervised exercises in patients with rotator cuff disease (stage II impingement syndrome): a prospective randomized, controlled study in 125 patients with a 2 ½ year follow-up. *J Shoulder Elbow Surg*. 1999;8:102–11.
- 12 Lombardi I, Magri A, Fleury A, da Silva A, Natour J. Progressive resistance training in patients with shoulder impingement syndrome: a randomized controlled trial. *Arth Rheum*. 2008;59:615–22.
- 13 Ludewig P, Borstad J. Effects of a home exercise programme on shoulder pain and functional status in construction workers. *Occup Environ Med*. 2003;60:841–9.
- 14 Walther M, Werner A, Stahlschmidt T, Woelfel R, Gohlke F. The subacromial impingement syndrome of the shoulder treated by conventional physiotherapy, self-training, and a shoulder brace: results of a prospective, randomized study. *J Shoulder Elbow Surg*. 2004;13:417–23.
- 15 Furlan D, Pennick V, Bombardier C, van Tulder M. 2009 Updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine*. 2009;34:1929–41.
- 16 van Tulder M, Suttrop M, Morton S, Bouter L, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. *Spine*. 2009;34:1685–92.
- 17 Dorrestijn O, Stevens M, Winters J, van der Meer K, Diercks R. Conservative or surgical treatment for subacromial impingement syndrome? A systematic review. *J Shoulder Elbow Surg*. 2009;18:652–60.
- 18 Kelly S, Wrightson P, Meads C. Clinical outcomes of exercise in the management of subacromial impingement syndrome: a systematic review. *Clin Rehabil*. 2010;24:99–109.
- 19 Kromer T, Tautenhahn U, de Bie R, Staal J, Bastiaenen C. Effects of physiotherapy in patients with shoulder impingement syndrome: a systematic review of the literature. *J Rehabil Med*. 2009;41:870–80.
- 20 Kuhn J. Exercise in the treatment of rotator cuff impingement: a systematic review and a synthesized evidence-based rehabilitation protocol. *J Shoulder Elbow Surg*. 2009;18:138–60.
- 21 Michener L, Walsworth M, Burnet E. Effectiveness of rehabilitation for patients with subacromial impingement syndrome: a systematic review. *J Hand Therapy*. 2004;17:152–64.
- 22 Trampas A, Kitsios A. Exercise and manual therapy for the treatment of impingement syndrome: a systematic review. *Phys Ther Rev*. 2006;11:125–42.
- 23 van der Heijden G, van der Windt D, de Winter A. Physiotherapy for patients with soft tissue disorders: a systematic review of randomised clinical trials. *BMJ*. 1997;315:25–30.